Offer #2021-03523

Post-Doctoral Research Visit F/M Emerging Paradigms for Resilient Computing

Renewable contract : Yes

Level of qualifications required : PhD or equivalent

Fonction : Post-Doctoral Research Visit

About the research centre or Inria department

The Inria Rennes - Bretagne Atlantique Centre is one of Inria's eight centres and has more than thirty research teams. The Inria Center 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 PMEs, large industrial groups, competitiveness clusters, research and higher education players, laboratories of excellence, technological research institute, etc.

Context

CAIRN (Energy-Efficient Computing Architectures) Inria project-team, Rennes, France.

https://team.inria.fr/cairn/

The CAIRN project-team pursuits cutting edge researches on new architectures, algorithms and design methods for flex- ible, secure, fault-tolerant, and energy-efficient domain-specific computing architectures. The team research directions are built around domain-specific systems in four main contexts: hardware accelerators, approximate computing, resilient computing, emerging technologies. The proposed research program perfectly fits in the CAIRN main research directions. The highly international research environment and the innovative studies conducted within CAIRN will provide all the favorable conditions for reinforcing existing international collaborations and developing new ones.

Assignment

Digital computing systems are an increasingly fundamental part of our existence. In the last decade, Artificial Intelligence (AI) has revolutionized the way people use and rely on computing machines [5–7]. The deep learning paradigm and, in particular, Deep Neural Networks (DNNs) enable electronic systems to perform increasingly complex tasks and are the focus of extensive studies, all over the world. To name a few, personal assistants help us to quickly retrieve information from the web and control smart objects in our houses, AI-based robots perform hard and repetitive tasks autonomously and help doctors with medical therapy, and AI-based algorithms help researchers in medical and drug research.

DNN-based systems have been studied from different standpoints in the last decades. First of all, the ability of a DNN to perform classification tasks with the highest possible accuracy has been the main focus for a long time [8]. Their incredibly high effectiveness comes at the price of an extremely large algorithmic complexity, thus of an enormous computational power.

Investigating new approaches to improve the efficiency of AI-based systems is necessary. From a general perspective, it is widely accepted that full-precision computing is inappropriate for next-generation AI-based applications due to hardware and energy costs. Therefore, the worldwide research community has moved towards the design of novel reduced-precision-computation techniques and low-cost hardware. Thus, novel non-conventional computing paradigms are increasingly employed to improve the energy efficiency of AI-based systems. Among others, popular research topics in this sense are neuromorphic computing and in-memory computing, which dispute the efficiency of conventional computing paradigms and promote a more brain-inspired methodology [12]. Another non-conventional computing paradigm, the Approximate Computing (AxC), has been increasingly applied to AI-based systems to push further their resource efficiency. AxC profits from the intrinsic error-tolerance of AI applications and systems – especially DNNs – to achieve high benefits in terms of resource efficiency [2,3]. Specifically, AxC carefully introduces some inaccuracy – that will be intrinsically tolerated – to increase the computing systems' efficiency in terms of performance, area, and power consumption.

Main activities

Therefore, this research positions mainly focuses on studying the Approximate Computing as an emerging paradigm to design the next-generation resource-efficient AI-based safety-critical systems. In details, in the context of AxC, there is the opportunity to relax the system reliability constraints to trade them off with important power savings and performance boosting [14, 15]. While this surely increases

system efficiency, on the other hand utilizing approximation techniques in safety-critical scenarios represents a delicate and tricky task. Indeed, despite the energy efficiency optimization opportunities, reliability still represents a key requirement in most advanced safety-critical computing system: sacrificing reliability could result in the production of more cost-efficient systems, but also in endangering human lives. In particular, approximate AI-based systems must be designed to be reliable in order to be used in safety- critical scenarios, such as autonomous driving, robotics, and e-health.

Keywords – Artificial Intelligence, Deep Learning, Safety-critical systems, Energy efficiency, Approximate Computing, Reliability, Testing, Diagnosis, Verification, Digital Design, Design-for-Reliability

Bibliography

[1] Q. Xu, et al. Approximate computing: A survey. IEEE Design Test, 33(1):8–22, Feb 2016.

[2] Sparsh Mittal. A survey of techniques for approximate computing. ACM Comput. Surv., 48(4):62:1–62:33, March 2016.

[3] V. K. Chippa, et al. Analysis and characterization of inherent application resilience for approximate computing. In 2013 50th ACM/EDAC/IEEE Design Automation Conference (DAC), pages 1–9, May 2013.

[4] Semeen Rehman, et al. Heterogeneous Approximate Multipliers: Architectures and Design Methodologies, pages 45–66. Springer International Publishing, 2019.

[5] Yann LeCun, et al. Deep learning. Nature, 521(7553):436–444, 2015.

[6] Cheng Jin, et al. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. Nature Communications, 11(1):5088, December 2020.

[7] David Silver, et al. Mastering the game of go with deep neural networks and tree search. Nature, 529(7587):484–489, 2016.

[8] Asifullah Khan, et al. A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review, 53(8):5455–5516, December 2020.

[9] Sparsh Mittal. A survey of fpga-based accelerators for convolutional neural networks. Neural Computing and Applications, Oct 2018.

[10] Maurizio Capra, et al. An Updated Survey of Efficient Hardware Architectures for Accelerating Deep Convolutional Neural Networks. Future Internet, 12(7):113, July 2020.

[11] Alfredo Canziani, et al. An analysis of deep neural network models for practical applications, 2017.

[12] C. Mead. Neuromorphic electronic systems. Proceedings of the IEEE, 78(10):1629–1636, October 1990.

[13] C. Torres-Huitzil and B. Girau. Fault and error tolerance in neural networks: A review. IEEE Access, 5:17322–17341, 2017.

[14] L. Anghel, et al. Test and Reliability in Approximate Computing. Journal of Electronic Testing, 34(4):375–387, August 2018.

[15] I. Polian. Test and reliability challenges for approximate circuitry. IEEE Embedded Systems Letters, 10(1):26–29, March 2018.

[16] M. Traiola, et al. Investigation of mean-error metrics for testing approximate integrated circuits. In 2018 IEEE In- ternational Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT), pages 1–6, Oct 2018.

[17] M. Traiola, et al. A test pattern generation technique for approximate circuits based on an ilp-formulated pattern selection procedure. IEEE Transactions on Nanotechnology, pages 1–1, 2019.

[18] M. Traiola, et al. Maximizing yield for approximate integrated circuits. In 2020 Design, Automation Test in Europe Conference Exhibition (DATE), pages 810–815, 2020.

[19] M. Traiola, et al. A survey of testing techniques for approximate integrated circuits. Proceedings of the IEEE, pages 1–17, 2020.

[20] M L. Bushnell and V D. Agrawal. Essentials of Electronic Testing for Digital, Memory, and Mixed-Signal VLSI Circuits. Springer, Boston, MA, 01 2000.

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)
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities

- Access to vocational training
- Social security coverage

Remuneration

Monthly gross salary amounting to 2653 euros.

General Information

- **Theme/Domain :** Architecture, Languages and Compilation
  System & Networks (BAP E)
- **Town/city :** Rennes
- **Inria Center :** Centre Inria de l'Université de Rennes
- **Starting date :** 2021-10-01
- **Duration of contract :** 1 year, 5 months
- **Deadline to apply :** 2021-05-31

Contacts

- **Inria Team :** CAIRN
- **Recruiter :**
  Sentieys Olivier / Olivier.Sentieys@irisa.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

Please submit online : your resume, cover letter and letters of recommendation eventually

For more information, please contact olivier.sentieys@inria.fr

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.