

Department	School of Computing
Supervisors	CHERUBIN, Stefano
Project Title	Approximate Computing for next-gen AI-specific Accelerators

PROJECT DESCRIPTION

Intelligent systems are pervading every aspect of modern society and their use has become essential in many industries, from big data to industry 4.0, from cloud computing to the Internet of Things, from smart cities to smart wearable devices.

They are deployed everywhere, from centralised supercomputers to pervasive fog and edge devices. The variety of these applications poses new research challenges in the field, most prominent the greenness of the artificial intelligence. Although one single artificial neural network will not pollute as much as a human person, the number of artificial brains each single person uses is rapidly increasing.

In the past years, the shape of intelligent computing systems has been twisted, melt, reforged and re-designed multiple times. Fog and liquid computing trends are enriching the computing scenario with new types of smart devices, each of them with unique capabilities and requirements. In this ever-changing world, great attention is to be reserved to the battery life and to the efficiency of the computing hardware. Manufacturers around the world are rushing to design the most efficient computers and supercomputers. The solution is almost always heterogeneity (Green500). Specialised pieces of hardware can improve the extra-functional properties of the computing system, *in primis* the CO2 footprint of the application they run. The most promising innovation that drives the evolution of the tensor architectures is their ability to move less data while delivering the same quality of results. Data transfer and data processing costs are on the critical path to the time- and energy-to-solution. Further developments are relaxing their requirements on computation accuracy to leverage the approximate computing paradigm.

Beyond the full implementation of the most common IEEE-754 standards lay a kaleidoscope of possibilities to implement, explore, and assess new data types which are often too complex to be manually engineered in existing applications.

In this project, you will work on automatic tools to analyse applications, transform their code, and improve their suitability to a given hardware accelerator. The successful candidate will explore the research questions that arise in use cases involving single-core, multi-core, many-core, and non-Von-Neumann computer architectures.

The engineering challenges in this project will contribute towards the improvement and adoption of open standard both in hardware and in software design, with SYCL and RISC-V as main technologies.

Prospective applicants are encouraged to contact the Supervisor before submitting their applications. Applications should make it clear the project you are applying for and the name of the supervisors.

Academic qualifications

A first degree (at least a 2.1) ideally in Computer Science or related discipline with a good fundamental knowledge of algorithms and data structures.

English language requirement

IELTS score must be at least 6.5 (with not less than 6.0 in each of the four components). Other, equivalent qualifications will be accepted. [Full details of the University's policy](#) are available online.

Essential attributes:

- Experience of fundamental C++ programming
- Competent in concurrent and parallel systems, hardware/software co-design, and code optimisations

- Knowledge of fundamentals of computer architectures
- Good written and oral communication skills
- Strong motivation, with evidence of independent research skills relevant to the project
- Good time management

Desirable attributes:

- Experience with one or more heterogeneity-aware code acceleration paradigms (OpenCL, CUDA, SYCL, etc.)
- Experience with large C++ code bases
- Experience with MLIR and/or LLVM framework
- Track record of open source code contributions

Indicative Bibliography	Cherubin, S., Agosta, G., Lasri, I., Rohou, E., & Sentieys, O. (2018). Implications of Reduced-Precision Computations in HPC: Performance, Energy and Error. In International Conference on Parallel Computing (ParCo) (pp. 297-306). Cherubin, S., & Agosta, G. (2020). Tools for reduced precision computation: a survey. ACM Computing Surveys (CSUR), 53(2), 1-35. (Green500) https://www.top500.org/lists/green500/
Enquiries	For informal enquiries about this PhD project, please contact s.cherubin@napier.ac.uk
Web page	https://www.napier.ac.uk/research-and-innovation/research-degrees/application-process