PhD Project Proposal

ECIT Interdisciplinary PhD Programme

Proposed Project Title: **Acceleration of k-mer counting in DNA sequencing using FPGA-based hardware**

Principal Supervisor(s): **Prof Roger Woods**
Dr Darragh McArt (CCRCB)

Project Description:

Counting the number of occurrences of every *k*-mer (substring of length *k*) is a central step in many analyses of DNA sequence in genomics research. In particular, the deep sequence coverage step needed in next-generation sequencing technologies has caused the amount of sequence to be processed during a genome project to grow rapidly. This has rendered current *k*-mer counting tools too slow and are deemed to be memory intensive. JELLYFISH is just one example of a tool for fast, memory-efficient counting of k-mers in DNA using an order of magnitude less memory and an order of magnitude faster than other k-mer counting packages. This is achieved by using an efficient encoding of a hash table and by exploiting the "compare-and-swap" CPU instruction to increase parallelism.

The purpose of this project is to look to accelerate this core *k*-mer counting function using FPGA technology. To this end, an initial exploration with Xilinx has led them to dedicate one of their hard work exploration cards, namely the Alveo Adaptable Accelerator Cards (https://www.xilinx.com/products/boards-and-kits/alveo.html) worth $10k to the University to allow this project to progress. The idea would be to create an accelerator technology of solutions to be developed to allow CCRCB to incorporate the technology and provide world-leading advances to be made by accelerating this core function.

Objectives:

- To work with CCRCB in order to produce accelerated solutions of the initial JELLYFISH algorithm for the core k-mer counting function with an advanced next-generation sequencing systems.
- To explore variants of algorithms for acceleration for *k*-mer counting using state-of-the-art FPGA-based hardware.
- To create a practical hardware/software acceleration system within CCRCB and explore extent of acceleration of core bioinformatics algorithms and their resulting performance capability.

Academic Requirements:

Students entering the programme will normally be required to have a 2.1 BSc/BEng in Computer Science, Electrical and Electronic Engineering, Bioinformatics, or a maths based engineering or physical science degree, or equivalent qualification recognised by the University. Students holding an appropriate MEng or MSc (Software conversion) will

normally be required to have a 2.1 or commendation (distinction) respectively. Furthermore, additional criteria may be applied. All applicants must have significant mathematical and programming experience.

GENERAL INFORMATION:
This 4 year PhD studentship, potentially funded by the Department for Employment and Learning (DEL), commences on 1 October 2019.

Eligibility for both fees and maintenance depends on the applicants being either an ordinary UK resident or those EU residents who have lived permanently in the UK for the 3 years immediately preceding the start of the studentship. Non UK residents who hold EU residency may also apply but if successful may receive fees only.

Applicants should apply electronically through the Queen's online application portal at: https://dap.qub.ac.uk/portal/

Deadline for applications: Friday 1 March 2019

Contact details:

Supervisor Name: Tel: +44 (0)28 9097
QUB Address: Email: