Approximate Computing for Energy-efficient Whole Genome Sequencing Systems

(NU+NTU) | Theme: Ageing and Digital Technologies

| Supervisory Team | Dr Rishad Shafik, Newcastle University, UK, www.rishadshafik.net
|                  | Dr Nachiket Kapre, Nanyang Technological University, Singapore, http://nachiket.github.io/
|                  | Dr Ian Wilson, Institute of Genetic Medicine, http://www.ncl.ac.uk/medicalsciences/research/groups/profile/ian.wilson
|                  | The Lead Supervisor is Early Career or newly hired Staff ☒

Key Words: Energy-efficiency, Translational technology, Genome sequencing, FPGA

Overview

Future generations of healthcare services are expected to be confronted by an overwhelming increase in ageing population. In the façade of such an increase, providing personalised diagnostics and drug delivery services is going to be extremely challenging. Biomedical applications have excellent potentials to address this challenge through translational technologies. A profound example of such an application is computational whole genome sequencing (WGS), which facilitates in-depth understanding of the microbial ecology and genetic basis of many complex human diseases and their traits. However, current WGS schemes are highly computationally intensive, as more than six billions of nucleotides are decompressed, filtered, sequenced with patterns and finally synthesised to extract useful information. As a result, these schemes have very low throughput and high energy consumptions; for example, one WGS can be carried out per 26 hours, incurring 112 mega Watt-hour energy consumption on a 1200-Watt supercomputer with 244 high-end processor cores [http://tinyurl.com/p3qesx9]. Such a low throughput and energy consumption may prove infeasible, when every newborn’s DNA will need to sequenced in the future to enable personalised diagnostics and drug delivery services. Indeed, to realise this important and promising translational technology, there are key research questions and challenges as follows:

1. Can the energy consumption be drastically reduced and the throughput be improved significantly?
2. Can such a technology be enabled at a much lower cost than the high-end supercomputing nodes?

This project will address the above research questions and set up a systematic design, and implementation of a low-cost, high-throughput and energy-efficient WGS electronic system. The throughput improvement and energy efficiency will be incorporated through approximate computing, which systematically incorporates scaling of the bit-width precision, such that computation and storage can be drastically reduced at the expense of deliberately introduced imprecisions. To design and implement the system, the project has the following specific objectives:

1. To study the relationships between WGS outcomes and the computation/precision scaling through sensitivity analyses.
2. To develop a systematic learning-based approach of bit-width precision and computation scaling to achieve throughput improvement and energy reductions, while also maintaining the required quality of WGS outcomes.
3. To design, implement and validate the approach on a high-performance and low-cost field-programmable gate array (FPGA) platform.

This project will benefit from complementary and cross-disciplinary collaboration between Newcastle University (NU) and Nanyang Technological University (NTU). The NU team benefits from research strength and partnership between adaptive electronic systems (Microsystems Research Group, MRG) and WGS algorithms (Institute of Genetic Medicine, IGM). The NTU Co-PI (Dr. Kapre) will lead the FPGA-based design, implementation and validation with his excellent knowledge and track record of the same. The project will benefit from industrial collaboration (eg. internship) with Altera and Microsoft (currently being discussed). Newcastle MRG is one of the leading groups in the UK for it’s research in energy-efficient electronic systems, while IGM is a key contributor of Genomics England’s 100,000 genomes flagship project and one of the leading Genomics Medicine centres in the UK. NTU is globally well-ranked and known for its excellent impact and quality of research.
Methodology

Fig. 1 shows the proposed WGS system design and implementation methodology using approximate computing, organised in the following three major tasks:

Task A: Study of Approximate Computing Opportunities: WGS outcomes are often significantly different from each other. For Mendelian (inheritance-based) diseases (such as diabetes), evolutionary WGS outcomes may be much simpler than those in Non-Mendelian (non-heritability-based) outcomes requiring diagnostics of diseases (such as cancer). For each outcome, the relationship between the quality of the WGS outcome and computational bit-width scaling will be extensively studied to find the opportunities of approximate computing. The relationship will be established through formulation of the outcome sensitivities for each data representation in the WGS algorithms. To carry out this study, a high-level WGS algorithm will be used in C/C++ and/or Matlab®.

Task B: Optimisation of Bit-width and Computation Scaling: Underpinning the study in Task A, an algorithmic approach will be investigated. The aim of such an approach will be to learn the trade-offs between bit-width/computation scaling and the WGS outcomes. For each WGS outcome, an optimal throughput improvement and energy reduction will be ensured, while also maintaining the required quality of outcome. Due to varying significance of outcomes and their different stages, the throughput improvement and energy reduction will be adapted and learnt during runtime for each WGS outcome. During this task, an OpenCL-enabled (C/C++) FPGA emulator will be used to develop the learning and optimisation algorithms.

Task C: FPGA-based Design, Implementation and Validation: The OpenCL design and the developed algorithms in Task B will be ported and implemented on an FPGA-based platform (with a number of connected Altera Stratix 10 systems). The implementation will then be comprehensively validated using a number of WGS input and outcome sets. The validations will be carried out using the samples from the IGM at Newcastle, and also from freely available databases from various other sources. The overall aim will be to demonstrate a significant throughput improvement of WGS at reduced energy consumption, implemented on a low-cost electronic system.

Timeline

<table>
<thead>
<tr>
<th>Task</th>
<th>Time</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>1-9m</td>
<td>Understand the WGS algorithm using C/C++/Matlab (sequencing, pruning, filtering and post-processing)</td>
</tr>
<tr>
<td>A2</td>
<td>7-20m</td>
<td>Study approximate computing opportunities in WGS</td>
</tr>
<tr>
<td>B1</td>
<td>18-24m</td>
<td>Formulate the relationships between WGS outcome and precision requirements</td>
</tr>
<tr>
<td>B2</td>
<td>22-34m</td>
<td>Develop learning and optimisation algorithms for throughput improvement and energy minimisation (C/C++/OpenCL)</td>
</tr>
<tr>
<td>C1</td>
<td>31-42m</td>
<td>Mapping WGS tasks on FPGA platform and pre-verification</td>
</tr>
<tr>
<td>C2</td>
<td>36-45m</td>
<td>Validation and case studies (IGM+external dataset)</td>
</tr>
<tr>
<td>D</td>
<td>13-45m</td>
<td>Publications (high-impact journals, dissertations, etc.)</td>
</tr>
</tbody>
</table>

References & Further Reading

Learning-based Design and Runtime Optimisation:
1. Soft error-aware design optimization of low power and time-constrained embedded systems. Shafik, R. et al. 2010. DATE.

FPGA-based Acceleration and Validations:

Genomics:

Further Information

For further details regarding this project, contact Dr Rishad Shafik: Merz Court E4.14, Newcastle University, Newcastle, NE1 7RU.
E-mail: Rishad.Shafik@newcastle.ac.uk.