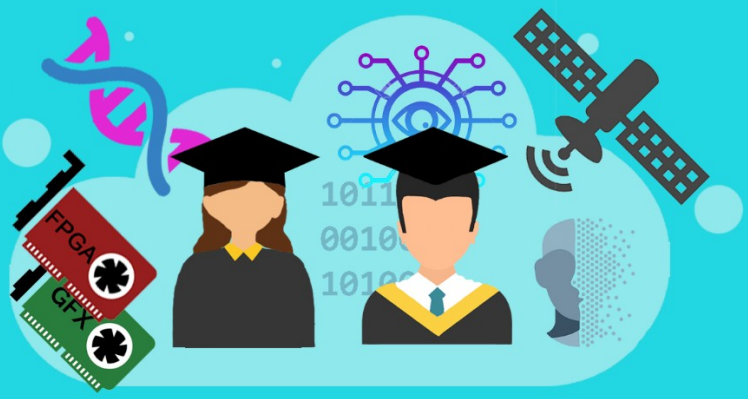
Diploma Thesis

Microproccessors and Digital Systems Laboratory



DATAFLOW ACCELERATION OF GENOME ANALYSIS WORKFLOWS FOR BIG DATA APPLICATIONS RUNNING ON A CLOUD INFRASTRUCTURE WITH HETEROGENEOUS ARCHITECTURE

The development of next-generation sequencing (NGS) technologies has dramatically changed the landscape of human genetics research. In a typical genomics analysis workflow, NGS platforms produce short fragments of DNA reads that are first aligned to a genome reference and then used for variant calling to identify differences and mutations. Many aligners and variant callers have been developed for that purpose and composed into diverse pipelines, utilized by research groups all over the world. However, the effective analysis of raw NGS data still poses a major challenge, since the ever-growing amount of sequencing data and high complexity of alignment and variant calling algorithms form bottlenecks in the standard genome analysis flow, that requires hours or even days to complete.

The scope of this diploma thesis is to contribute in building a library of accelerators, that optimize the genome analysis workflows. Preconfigured workflows available in popular third-party tool Seqmule will be extensively profiled in order to identify suitable candidates for acceleration focusing on the variant calling stage. The most suitable variant caller will be carefully inspected to pinpoint its major bottleneck. This algorithmic segment will be implemented on an FPGA platform, forming an accelerated SW/HW co-designed solution. The hardware implementation will be realized utilizing the dataflow computing model proposed by Maxeler Technologies. Maxeler Technologies provides a special platform that includes 8 FPGA connected devices that communicate with the host machine via Infiniband. The resulted design will be incorporated in an existing framework that integrates cloud technologies in a distributed system with Maxeler heterogeneous architecture. The execution of the accelerated workflow will be invoked in a docker container environment, whose deployment, scaling and operation is managed by Kubernetes Resource manager.

CONTACT INFORMATION: Prof. Dimitrios Soudris: dsoudris@microlab.ntua.gr, Konstantina Koliogeorgi: konstantina@microlab.ntua.gr, Sotirios Xydis: sxydis@microlab.ntua.gr