

Sequence Similarity Search on Reconfigurable Computing System

Sajish Chandrababu¹, Yogindra Abhyankar¹ and Rajendra Joshi²

¹Hardware Technology Development Group, ²Scientific and Engineering Computing Group

Centre for Development of Advanced Computing

Pune, India

E-mail: {yogindra,rajendra}@cdac.in

Abstract

Genome sequencing has been one of the great achievements of biotechnology. The future of bioinformatics lies in the analysis of the vast ocean of data generated by whole genome sequencing. There are tremendous challenges in building a high throughput environment for whole genome analysis of multiple organisms. The Smith-Waterman algorithm provides sequence alignments with high sensitivity but suffers from high computational requirement. We report an accelerated Smith-Waterman solution using FPGA based Reconfigurable Computing (RC) that significantly brings down the search time. Our solution handles large lengths up to 1 M characters each for the query as well as database sequences and do not restrict on using a particular brand of processor. The experimental benchmarks show 90X speedup in the execution time against a purely software solution. Reported performance figures include cases, where reduction in search time from 4 days to merely 59 minutes is obtained. Multiple queries with an estimated search time of 74 days were completed in just 20 hours by our RC solution.