Accelerating Genome Sequence Alignment on a RISC-V Many-Core Cluster

Group No: 38

Jitesh Kumar Nayak PES2UG20EC057

Abhiram Gopal Dasika PES2UG20EC003

Neha C Waghmore PES2UG20EC092

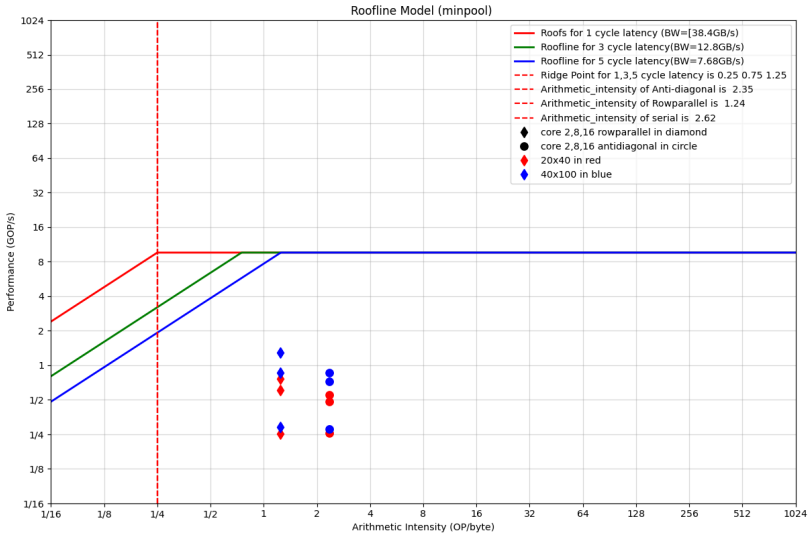
Guide: Dr Madhura Puraprajna

ABSTRACT

Read mapping is a critical and computationally demanding step in genome sequencing, with sequence alignment constituting the most time-consuming portion. This project aims to significantly accelerate this process by parallelizing the implementation of the Smith-Waterman algorithm, the most widely used approach for sequence alignment, which relies on a dynamic programming (DP) matrix to efficiently compute optimal alignments. To achieve this, we will leverage the MemPool architecture, a scalable many-core platform distinguished by its low-latency shared memory

Two distinct parallelization strategies are explored: The anti-diagonal method parallelizes computations along the natural diagonal data dependency of the dynamic programming matrix, while the row-parallel method focuses on concurrent row calculations within the dynamic programming matrix by modifying the smith-waterman algorithm to make data

For a 40-character query and 100-character reference, the anti-diagonal method exhibited limited scalability. Due to the presence of fewer elements in initial and final diagonals, many cores remained idle, resulting in modest speedups (1.66x, 2.79x, 4.35x, and 5.4x) for core-count 2,4,8,16. In contrast, row-parallelism demonstrated strong scalability, achieving significant speedups (1.95x, 3.47x, 5.78x, and 8.66x) for core-count 2,4,8,16 by effectively utilizing many-core resources. This approach, while leading to a slight reduction in arithmetic intensity, achieved higher throughput than the anti-diagonal method while remaining within the compute-bound region.



ROOFLINE MODEL