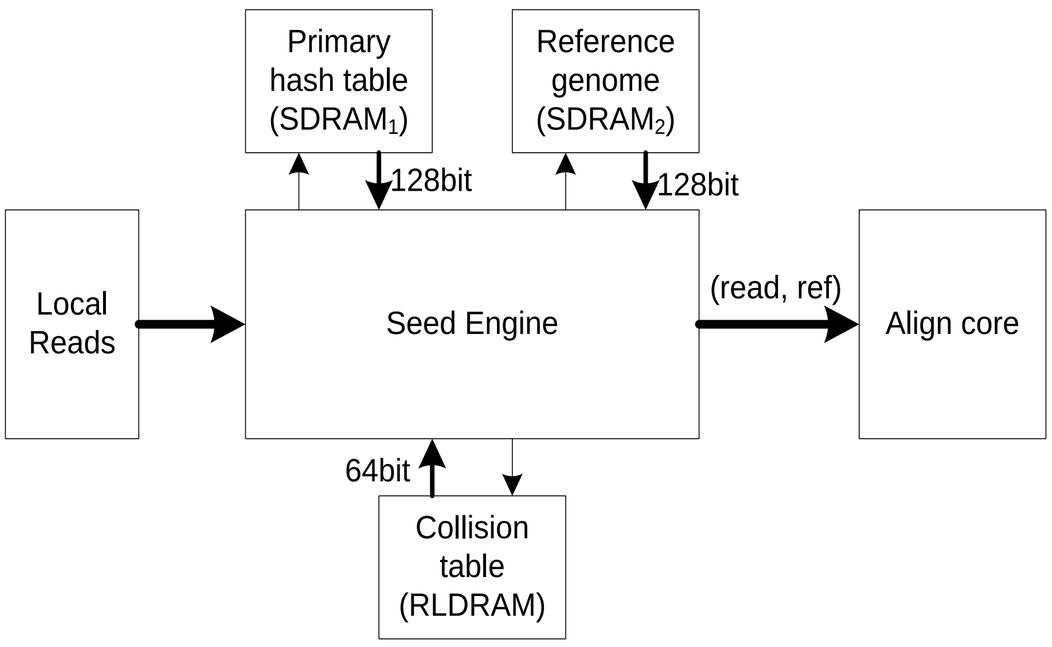
NANYANG TECHNOLOGICAL UNIVERSITY

Centre for High Performance Embedded Systems

Presented by:
Mr. Alok Prakash

An FPGA aligner for short read mapping

The explosive growth of short read datasets produced by high throughput DNA sequencing tools poses a challenge to the mapping of short reads to a reference genome in terms of sensitivity and execution speed. In this design, we use the "seed and extend" strategy to map the short reads in two steps: 1) find the short exact matches (i.e. candidates) between the read and the reference genome; 2) extend the matches using a more compute intensive semi-global alignment algorithm. Both of the steps are implemented on a Xilinx LX330 chip to shorten the computation time.



The final system achieves over 6 times speedup against GASSST with a similar sensitivity; around 5 times speedup against BWA with a higher sensitivity using one million simulated short read datasets. Currently the performance is limited by the frequent off-chip memory access within the seed generation step.

To better utilizing FPGAs' outstanding performance on fine-grained pipelining and parallelism, we have proposed a parallel banded semi-global alignment architecture to approximate the conventional algorithm. The search space is divided into multiple blocks; different blocks can be computed in parallel and the final alignment score is computed using a tree structure to further reduce the computation time.