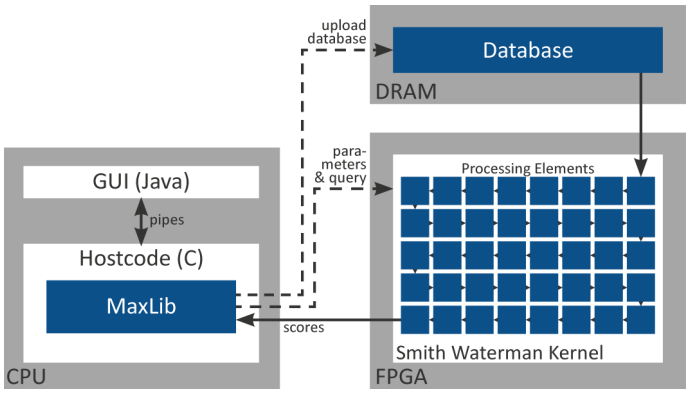
Smith Waterman Demo

Sequence Alignment for Nucleotides and Proteins

At a Glance

- Amino acid pairwise alignment using the Smith-Waterman algorithm and a linear gap model.
- BLOSUM62 scoring matrix.
- Variable opening gap penalty.
- 800 GCUPS on a single 1U node.
- Accelerator code written in MaxCompiler, hostcode written in C, GUI in Java. C and Java code communicate over Unix pipes.

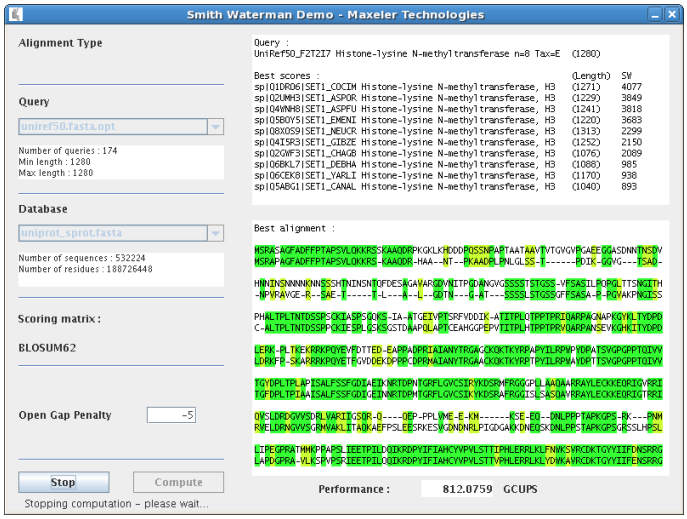


Accelerated application architecture

The Smith-Waterman algorithm is a dynamic programming technique used in bioinformatics to perform similarity searches between DNA or protein sequences. It is the most accurate algorithm for this purpose, but also the most compute-intensive with a quadratic complexity, and tends to be more and more time consuming with the exponential growth of protein libraries. Maxeler's Smith-Waterman application demonstrates an accelerated solution applied to bioinformatics for pairwise local alignments.

This application performs local alignments of amino acid sequences against a database using a linear gap model. The database is loaded into the on-board DRAM memory before the computation, while each query sequence is held in on-chip memory for fast repeated access. Performing several queries against the same database is achieved by uploading the sequences one after another onto the FPGA. The database is streamed through the processing elements during the computation and the scores for each alignment are streamed out to the host via the PCI Express bus. The best scores are sent to the GUI, where backtracking is computed in order to display the best alignment.

Maxeler's implementation of Smith-Waterman running on a single 1U node achieves a 16x speed-up over a 12-core conventional node running highly optimized FASTA SSEARCH. Maxeler's framework allows a great flexibility to satisfy various needs: customized hardware designs with a different gap model or performing nucleotide alignments can be created very easily with MaxCompiler.



Application GUI

How to Run the Demo

- Run './run.sh' in the demo directory to launch the GUI.
- Select a query file and a database file in FASTA format. The query file can contain several sequences. In this case, each sequence will be compared to the sequences in the database one after the other.
- Enter an opening gap penalty.
- Click 'Compute' to start the alignment using the BLOSUM62 scoring matrix.
- The best ten results are displayed in the top right frame. The aligned part of the best match is displayed in the bottom right frame. Matching amino acids are highlighted in green, similarities in yellow. The throughput is displayed below in GCUPS (billions of Cell Update per Second)

Maxeler Advantage

- Fast and high-capacity on-card DRAM to store large databases.
- A single Maxeler 1U node containing 4 MAX3 cards provides the same performance as 16 conventional 12-core CPU nodes running SSE2 optimized Smith-Waterman local alignment.

MAXELER Technologies
MAXIMUM PERFORMANCE COMPUTING