High Performance Smith-Waterman Local Sequence Alignment

This is a summary of a presentation at the Symposium on Verification at the University of Tokyo, Japan, 2 February, 2011.

About the Authors

Akira Fukui is a graduate student in the Fujita Lab at the University of Tokyo.

Summary

We present an dataflow engine-based acceleration of the Smith-Waterman algorithm for local sequence alignment. It is used for similarity searching in bioinformatics protein sequence databases, although often less precise tools, such as BLAST, are preferred because of its high computational cost.

This presentation reports a Maxeler acceleration project which achieves 240,000 MCUPS (million cell updates per second) in a 1U server, representing a 128x speedup over the industry-standard SSE-accelerated SSEARCH implementation (8-core Intel X5570 2.93GHz Nehalem EP), using a Maxeler MaxNode with 4 MAX2 DFEs.

As illustrated in Figure 1, Smith-Waterman has a wavefront dependency structure, in which parallelism is present along the diagonals of the matrix. To implement this efficiently, our array of cell update processing units (PU in Figure 2) operates on a stream of independent tasks in order to exploit deep pipeline parallelism despite the tight wavefront dependency structure, in which parallelism is present along the diagonals of the matrix (see Figure 3). This loop-tiling optimization, sometimes called c-slowing, is analogous to simultaneous multithreading (SMT) in microprocessors and enables full util-

Figure 1: Comparison of two sequences S and T, requires a cell update that depends on left, upper and diagonal neighbours.

Figure 2: The dependence structure limits available parallelism.

Figure 3: Interleaved scheduling of matching tasks on the processor array.

Original Publication Date: 2 February 2011

www.maxeler.com