Accelerating Genome Sequencing 100X with FPGAs

Olaf O. Storaasli
Future Technologies Group, Oak Ridge National Laboratory
Oak Ridge TN 37831 Olaf@ornl.gov

Abstract

The performance of two Cray XD1 systems with Virtex-II Pro 50 and Virtex-4 LX160 FPGAs was evaluated using the FASTA computational biology program for human genome (DNA and protein) sequence comparisons. Scalable FPGA speedups of 50X (Virtex-II Pro 50) and 100X (Virtex-4 LX160) over a 2.2 GHz Opteron were obtained. FPGA coding issues for human genome data are described.

FASTA Algorithm

FASTA [**1**] is used for protein: protein, DNA:DNA, protein: translated DNA and ordered or unordered peptide searches. It calculates similarity statistics for biologists to determine if alignments are random or homotopic. FASTA, who's input format is widely used by other search tools (i.e. BLAST [**2**]) relies on ssearch34 code. 98.6% of search34 time is spent in the FLOCAL_ALIGN function, a Smith-Waterman FPGA pipeline algorithm [**3-5**] (**Fig. 1**) to calculate the maximum alignment score for two sequences.

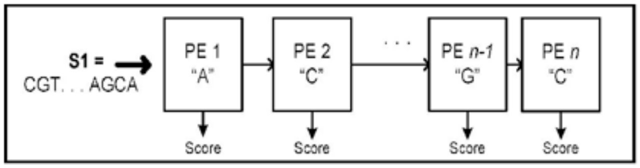

Figure 1. Smith-Waterman FPGA Pipeline

One query character is preloaded into each processing element which then calculate scores in the column of that query character. The database string (S1) is shifted through the pipeline so each database character is compared to each query character. A resulting table of scores results from computing scores in parallel.

OpenFPGA Benchmark Results:

Results were obtained for the comprehensive 4GB human genome sequencing openfpga.org benchmark.

Bacillus_anthracis DNA comparison: Genome matching was performed on Virtex2, Virtex4 and Opetron Cray XD1 configurations for 18 DNA query sequences: AE017024-AE017041 on a large database, AE016879 for two outputs:

Detailed: -Q –H –f -10 –g -3 **–d 10** –b 10 –s
Minimal: -Q –H –f -10 –g -3 **–d 0** –b 10 –s

Each query sequence (~300 thousand characters) was compared with the 5 million character database, runs which took over 3 days on the 2.2 GHz Opteron. As the FPGA Smith-Waterman code was limited to a maximum query size of 16k characters (and maximum database size of 512k characters), code was written to split the input query and database into smaller sequences. Search34 results were then obtained for 16k and 8k query sizes for two output options on two Cray XD1 systems (ORNL's Tiger-Virtex-II Pro 50 and Cray's Pacific-Virtex4 LX160) and compared with Opteron to determine FPGA speedups (**Figs. 2-3**).

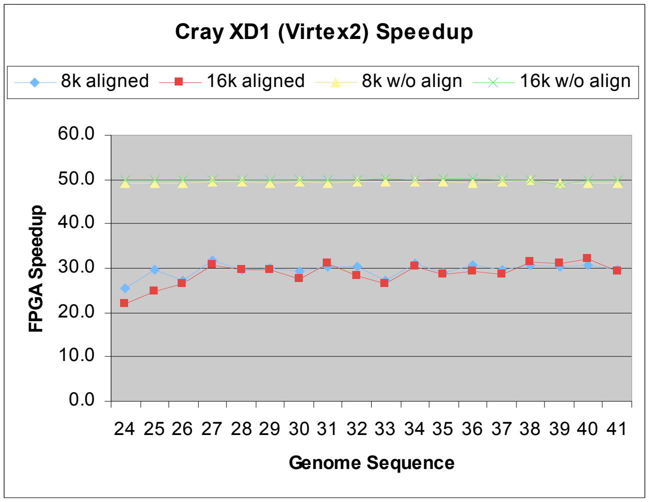

Figure 2. Virtex-II Pro 50 FPGA speedup

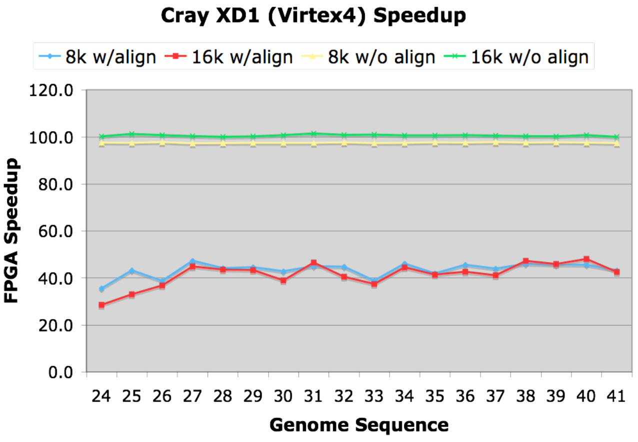

Figure 3. Virtex-4 LX160 speedup

Figsures 2-3 show that generating detailed alignment sequences results in FPGA speedups (blue and red curves) of 29X and 43X on Virtex2 and Viretx4 (with standard deviations of 1.8 and 3.9), respectively. IO, performed by the Opteron, was not optimized as it was a minor part of the Opteron's 75 hour execution time. However, for the FPGA code, generating this additional output slowed the code from 50X and 100X speedups (with standard deviations of 0.16 and 0.13) obtained for minimal output. 16k query sizes produced only slightly better performance than those for 8k.

Disclaimers and footnotes.

Clearly, the Virtex4 FPGA's 100X speedup outperformed Virtex2's 50X speedup, so that searches formerly taking 100 days (ie: 14 weeks) can now be completed in one day.

Analyis of results: The Virtex-4 LX160 with about 3x more logic area than the Virtex-II Pro 50, can process more copies of the algorithm in parallel. This additional logic space means it can run faster even though it runs at a slightly slower clock speed (125MHz) compared to the Virtex-II Pro 50 (140 MHz). The Virtex-4 (LX160) design has 128 SWPEs, compared to 48 for the Virtex-II Pro 50. As the timings indicate, it does more work, but it's clock speed is slightly less as it has more silicon area taking signals longer to traverse across the FPGA. More code optimization is possible which could speed the FPGA Smith-Waterman code by another factor of two. Such optimization promises potential speedups of 200X. FPGA designs generally run at different clock frequencies, however, there is a maximum clock speed for a given FPGA which a given design will run slower than unless it is very simple or extremely well written. The original SWA design frequency started out at less than 100 MHz on the Virtex-II Pro 50 and was increased to 140 MHz by optimizing the design. Similar improvements are also possible for the Virtex-4 LX160 code.

Query and Database Sizes: As little speedup difference was observed for 8k and 16k query sequence lengths, additional runs were made to determine the impact of query and database size on FPGA speedup. The first query sequence and database set was run an additional 30 times splitting the query sizes into sequences of length 1k, 2k, 4k, 8k, and 16k. Then, the database was split into sequences of sizes 16k, 32k, 64k, 128k, 256k, and 512k characters. FPGA speedups were then obtained by comparing the Virtex-II Pro 50 times to those for the Opteron (**Fig. 4**).

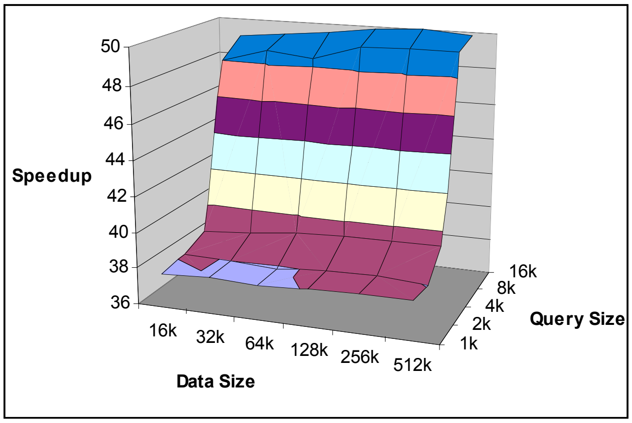

Figure 4. Speedup for Virtex-II Pro 50 FPGA

Fig. 4 indicates that database size has little impact on the FPGA algorithm speedup. Query sequence sizes of 8k to 16k both result in excellent speedups near 50X (similar to **Fig. 2**) while smaller query sizes of 1k still give respectable speedups of 37X. As before, one can expect speedups of 100X for longer query sizes on the Virtex4.

Amino Acid Search: FPGA speedups (Fig. 4) for Amino acid queries (myc, ras, src) in the openfpga.org benchmark showed similar results but with a wider variation among chromosomes (particularly src) which is,attributed to longer sequence lengths.

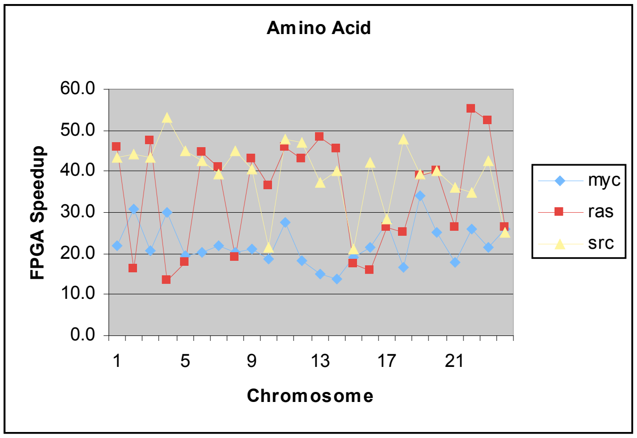

Figure 5. Virtex-II speedup for myc, ras and src sequences

Conclusions

FPGA performance was evaluated using the FASTA code for comprehensive biological DNA and amino acid sequencing on Cray XD1 computers with both Virtex-II Pro 50 and Virtex-4 LX160 FPGAs. Significant speedups of up to 100x over 2.2 GHz Opteron processors were observed. These results indicate similar speedups are likely for acceleration modules (DRC's and Xtreme data) that fit in Opteron sockets, whether in small embedded systems or Cray XT supercomputers.

References

[1] FASTA Sequence Comparison Program, fasta.bioch.virginia.edu

[2] MitrionC/BLAST, http://www.hpcwire.com/hpc/1274236.html

[3]Yim, Michael, Jacobs, Adam and George, Alan, Performance evaluation of the Cray Bioscience Applications Package on the XD1 (Cray White Paper).

[4]Margerm, Steve, and Maltby, Jim; Accelerating the Smith-Waterman Algorithm on the Cray XD1 (Cray White Paper WP-0060406) 2006.

[5]Storaasli, Olaf, Yu, Weikuan, Strenski, Dave, and Malby, Jim; Perfomance Evaluation of FPGA-Based Biological Applications, Cray Users Group Proceedings, Seattle WA, May 2007.