

Accelerating Genome Sequencing 100-1000X with FPGAs

Olaf O. Storaasli, Oak Ridge National Laboratory, Olaf@ornl.gov 865-574-0494
Future Technologies Group, MS-6173 Oak Ridge TN 37831-6173
Dave Strenski, Cray, Inc., Stren@cray.com

Abstract: The performance of FPGAs on Cray XD1 and SGI/RASC systems (with Virtex-II Pro 50 and Virtex-4 LX160 FPGAs) was evaluated for human genome sequencing using FASTA¹ and BLAST², respectively. Scalable speedups of 100X for a Virtex-4 FPGA (and exceeding 1000X for multiple FPGAs) over a 2.2 GHz Opteron were achieved. Similar FPGA speedups were achieved using BLAST on a SGI/RASC system at Oak Ridge National Laboratory.

Genome Sequencing Algorithm: Cray XD1: The maximum alignment scores for two DNA sequences were computed using FASTA with the VHDL Smith-Waterman pipeline algorithm³⁻⁵ taking 98.6% of the overall computation time (Fig. 1).

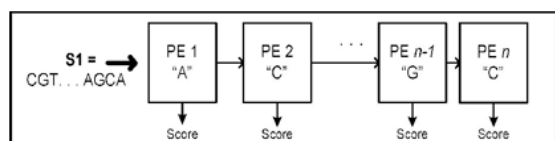


Figure 1. Smith-Waterman FPGA Pipeline

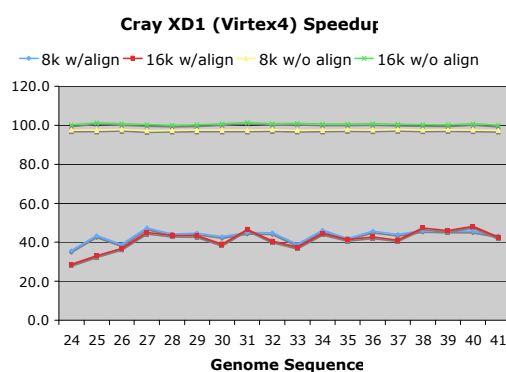


Figure 2. Virtex-4 LX160 speedup

A query character is preloaded in each processing element, which then calculates scores in the column of that query character. Each database string character (S1) is shifted through the pipeline and compared in parallel to each query character resulting in a table of scores. Speedups were obtained for from 1 to 150 FPGAs as compared with one 2.2 GHz Opteron.

SGI/RASC: Alignment scores for two DNA sequences were likewise computed using BLAST. Speedup results for 1 and 2 FPGAs were obtained and compared to a single Xeon processor.

Results: *Cray XD1:* The *Bacillus anthracis* openfpga.org benchmark with 18 DNA query sequences: AE017024- AE017041 (x-axis in Fig 2) of a 4 GB human genome database, AE016879. Each 300K character query sequence was compared with the 5 million character database, runs which exceeded 3 days on the 2.2 GHz Opteron. Speedups of 100X more than an Opteron (Fig. 2), for both 16k and 8k query sizes, were achieved on a Cray XD1-Virtex4 LX160 system (for minimal alignment output). This speedup enables 100-day searches to be completed in one day at twice the speedup obtained on Virtex II FPGAs used in ORNL and Naval Research Laboratory's (NRL) Cray XD1s. However, speedups on ORNL and NRLs Virtex II equipped Cray XD1's with 6 and 150 FPGAs, respectively show remarkable scalability, nearly proportional to the number of FPGAs used. The general trend shown in Fig 3 for Virtex II FPGA's is that Speedup increases rapidly with increased query size and even shows a slight increase with data size.

SGI/RASC: Similar speedups were realized using the initial Mitrion BLAST code at ORNL with two Virtex-4 FPGAs. Even better results are expected, and will be presented for the new Mitrion BLAST to be released in mid-February.

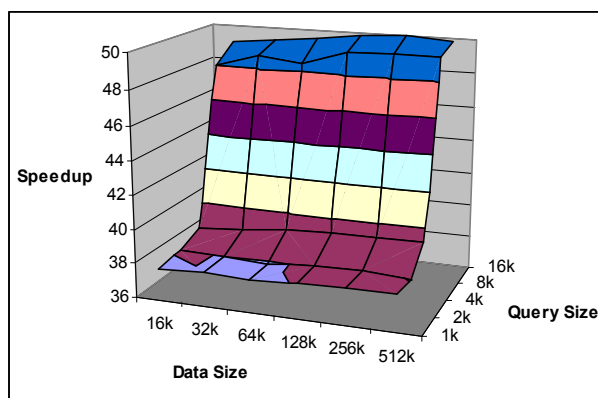


Figure 3. Speedup for Virtex-II Pro 50 FPGA

Additional results showing details of a myriad of multi-FPGA results on the NRL Cray XD1 with 150 FPGAs will also be presented at MRSC08.

Analysis of results: The 125MHz Virtex-4 (LX160) clock speed allows 128 SWPEs. As the timings indicate, it does more work despite its slower clock speed. The Cray XD1 VHDL code is not optimal and could perhaps achieve another factor of two speedup. Just as the Smith Waterman code began with a design frequency less than 100 MHz on the Virtex-II, and was optimized to 140 MHz, similar speedup is also possible for the Virtex-4 LX160 code.

Conclusions: FPGA performance was evaluated using FASTA/BLAST codes for DNA sequencing on Cray XD1/SGI-RASC computers. Significant speedups of up to 100x over 2.2 GHz Opteron processors were observed with near-scalable speedup in proportion to the number of FPGAs realized for up to 150 FPGAs used. These results indicate similar speedups are likely for acceleration modules (DRC's and XtremeData) 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; Performance Evaluation of FPGA-Based Biological Applications, Cray Users Group Proceedings, Seattle WA, May 2007.

Acknowledgements

1. Research sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle for the U. S. Department of Energy under Contract No. DE-AC05-00OR22725.
2. The authors wish to thank the U. S. Naval Research Laboratory for the use of their Cray XD1 supercomputer equipped with 150 Virtex II FPGAs and additional Virtex-4 FPGAs.