

Xia, Fei; Dou, Yong; Zhou, Dan; Li, Xin

Fine-grained parallel RNA secondary structure prediction using SCFGs on FPGA. (English)

Zbl 1204.68269

Parallel Comput. 36, No. 9, 516-530 (2010).

Summary: In the field of RNA secondary structure prediction, the CYK (Coche-Younger-Kasami) algorithm is one of the most popular methods using a SCFG (stochastic context-free grammar) model. Accelerating SCFGs for large models and large RNA database searching becomes a challenging task in computational bioinformatics because the parallel efficiency of general purpose computer systems is limited by the $O(L^3)$ computational complexity and by complicated data dependences. Furthermore, large scale parallel computers are too expensive to be easily accessible to many research institutes. Recently, FPGA chips have emerged as one promising application accelerator to accelerate the CYK algorithm by exploiting a fine-grained custom design. We propose a systolic-like array structure including one master PE and multiple slave PEs for the fine-grained hardware implementation on FPGA to accelerate the CYK/inside algorithm with Query-Dependent Banding (QDB) heuristics. We partition the tasks by columns and assign them to PEs for load balance. We exploit data reuse schemes to reduce the need to load matrices from external memory. The experimental results show a speedup factor of more than 14x over the Infernal - 1.0 with QDB optimization for the alignment of a single long RNA sequence to a large CM model with thousands of states running on a PC platform with Intel Dual-core 2.5 GHz CPU. The computational power of our accelerator is comparable to that of a PC cluster consisting of 16 Intel-Xeon 2.0 GHz Quad CPUs for large-scale database alignment applications (cmsearch) with multiple input sequences, but the power consumption is only about 10% of that of the cluster.

MSC:

68W10 Parallel algorithms in computer science

68M99 Computer system organization

Keywords:

bioinformatics; RNA; secondary structure prediction; SCFG model; parallel CYK algorithm; FPGA; hardware accelerator

Software:

BLAST ; ClustalW; Infernal; PSI-BLAST

Full Text: [DOI](#)

References:

- [1] Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic acids research* 25, 3389-3402 (1997)
- [2] W.R. Pearson, D.J. Lipman, Improved tools for biological sequence comparison, in: *Proceedings of the National Academy of Sciences*, 85, USA 1988, pp. 2444 – 2448.
- [3] (2009) Database Searching Tool: HMMER[Online]. Available: <http://hmmer.janelia.org/>
- [4] Thompson, J. D.; Higgins, D. G.; Gibson, T. J.: CLUSTAL W. Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice, *Nucleic acids research* 22, 4673-4680 (1994)
- [5] Zuker, M.; Stiegler, Patrick: Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information, *Nucleic acids research* 9, 133-148 (1981)
- [6] Gutell, R. R.; Power, A.; Hertz, G. Z.; Putz, E. J.; Stormo, G. D.: Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods, *Nucleic acids research* 20, 5785-5795 (1992)
- [7] Durbin, R.; Eddy, S. R.; Krogh, A.; Mitchison, G. J.: *Biological sequence analysis: probabilistic models of proteins and nucleic acids*, (1998) · [Zbl 0929.92010](#)

- [8] Rivas, E.; Eddy, S. R.: A dynamic programming algorithm for RNA structure prediction including pseudoknots, *Journal of molecular biology* 285, 2053-2068 (1999)
- [9] Eddy, Sean R.: What is dynamic programming?, *Journal of nature biotechnology* 22, No. 7 (2004)
- [10] Eddy, Sean R.: A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure, *BMC bioinformatics* 3, 18 (2002)
- [11] Liu, Tong; Schmidt, Bertil: Parallel RNA secondary structure prediction using stochastic context-free grammars, *Concurrency and computation: practice and experience* 17, 1669-1685 (2005)
- [12] Jane C. Hill and Andrew Wayne, A CYK approach to parsing in parallel: a case study, in: *Technical Symposium on Computer Science Education, Proceedings of the twenty-second SIGCSE Technical Symposium on Computer science education*, San Antonio, Texas, USA, 1991.
- [13] Guangming Tan, Shengzhong Feng, and Ninghui Sun, Exploiting parallelization for RNA secondary structure prediction in cluster. in *Proceedings of 5th International Conference on Computational Science (ICCS'05)*, LNCS 3516, Atlanta, GA, USA, 2005, pp. 979 – 982. · [Zbl 1120.92300](#)
- [14] G. Tan, L. Xu, S. Feng, and N. Sun, An experimental study of optimizing bioinformatics applications, in *Proceedings of IEEE International Parallel and Distributed Processing Symposium (IPDPS'06)*, Rhodes Island, Greece, 2006, pp. 25 – 29.
- [15] A. Jacob, J. Buhler, et al. Accelerating nussinov RNA secondary structure prediction with systolic arrays on FPGAs, in *Proceedings of IEEE International Conference on Application-specific Systems, Architectures and Processors (ASAP'08)*, Leuven, Belgium, 2008, pp. 191 – 196.
- [16] Cristian Ciressan, Eduardo Sanchez, Martin Rajman, and Jean-Cedric Chappelier, An FPGA-based coprocessor for the parsing of context-free grammars, in *Proceedings of IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM'2000)*, Napa, CA, USA, April 2000, pp. 236 – 245. · [Zbl 1001.68840](#)
- [17] Cristian Ciressan, Eduardo Sanchez, Martin Rajman, and Jean-Cedric Chappelier, An FPGA-based syntactic parser for real-life almost unrestricted context-free grammars, in *Proceedings of IEEE International Conference on Field Programmable Logic and Application (FPL'2001)*, LNCS 2147, Lausanne, Switzerland, October 2001, pp. 590 – 594. · [Zbl 1001.68840](#) · link.springer.de
- [18] Eddy, Sean R.; Durbin, Richard: RNA sequence analysis using covariance models, *Nucleic acids research* 22, No. 11, 2079-2088 (1994)
- [19] Lari, K.; Young, S. J.: Applications of stochastic context-free grammars using the inside – outside algorithm, *Computer speech and languages* 5, 237-257 (1991)
- [20] (2009) CPU Power Dissipation Statistic[Online]. Available: http://en.wikipedia.org/wiki/CPU_power_dissipation>.
- [21] Nussinov, R.; Pieczenik, G.; Griggs, J. R.; Kleitman, D.: Algorithms for loop matchings, *SIAM journal on applied mathematics* 35, 68-82 (1978) · [Zbl 0411.92008](#) · [doi:10.1137/0135006](https://doi.org/10.1137/0135006)
- [22] Nawrocki, E. P.; Kolbe, D. L.; Eddy, S. R.: Infernal 1.0: inference of RNA alignments, *Bioinformatics* 25, No. 10, 1335-1337 (2009)
- [23] J. Moscola, R. Cytron, Y. Cho, Hardware-accelerated RNA secondary-structure alignment, in: *Transactions on Reconfigurable Technology and Systems (TRETTS)*, September 2010, in press.
- [24] Yong Dou, Fei Xia, and Jingfei Jiang. Fine-grained parallel application specific computing for RNA secondary structure prediction using SCFGS on FPGA, in *Proceedings of ACM/IEEE International Conference on Compilers, Architecture and Synthesis for Embedded Systems archive (CASES'2009)*, Grenoble, France October 2009, pp. 107 – 116.
- [25] Nawrocki, Eric P.; Eddy, Sean R.: Query-dependent banding (QDB) for faster RNA similarity searches, *Plos computational biology* 3, No. 3, 0540-0554 (2007)
- [26] Yong Dou, Fei Xia, Xingming Zhou, and Xuejun Yang, Fine-grained parallel application specific computing for RNA secondary structure prediction on FPGA, in *Proceedings of IEEE 26th International Conference on Computer Design (ICCD'2008)*, Lake Tahoe, California, USA, October 2008, pp. 240 – 247.

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.