FPGA implementation of a greedy scheme for Bioinformatics applications

Rampone S (1), Aloisio A (2), Izzo V (2)

(1) Universita' del Sannio - Research Centre on Software Technology RCOST - Benevento, Italy
(2) Universita' di Napoli "Federico II" - Dipartimento di Scienze Fisiche and INFN - Napoli, Italy

Motivation
In the past decade there has been an explosive growth of biological data, including genome projects, proteomics, protein structure determination, cellular regulatory mechanisms, and the rapid expansion in digitization of patient biological data. Although raw computational power follows "Moore's Law", the genomic data at GenBank (the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences) is doubling every six months. Proteomic and cellular imaging data appear to grow even faster. Post-genomic-era bioinformatics will require high-performance computing power of the order of several hundreds of teraflops or more. Moreover a growing number of related problems is posed as complex optimization.

Methods
In recent years, FPGAs, short for Field-Programmable Gate Arrays, logic programmable chips [1] [2] have emerged as high-performance computing accelerators capable of implementing finegrained, massively parallelized versions of computationally intensive algorithms [3]. In particular several problems arising in biomedical and bioinformatics research can be viewed as finding the optimal covering of a finite set [4] [5]. While the Set Covering problem is known to be NPcomplete [6] a number of approximation heuristics have been proposed. The most efficient schema remains the greedy one [7]. Recently, a new greedy algorithm for approximating minimum set cover has been presented [8]. The algorithm, while not randomized, is based on a probability distribution that leads the greedy choice. It shows very good empirical performances and it has successfully been applied in wireless network applications [9] [10]. While efficient implementations are given, the cost of probability distribution evaluation can still be unaffordable in massive realtime applications. In this paper we describe an implementation based on a FPGA of a tailored version of the algorithm. It makes the algorithm suitable for several real world bioinformatics problems.

Results
The test results show very good empirical performances on the used benchmarks. The speed up of our approach is also successfully tested.

Contact email: rampone@unisannio.it

References
1. B. L. Hutchings, and M. J. Wirthlin, "Implementation approaches for reconfigurable logic applications", inW. Moore andW. Luk, editors Field- Programmable Logic and Applications, Oxford, Springer Verlag, 1995, pp. 419-428.
2. W. Mangione-Smith, B. Hutchings, D. Andrews, A. DeHone, C. Ebeling, R. Hartenstein, O. Mencer, J. Morris, V. Prasanna, and H. Spaanenburg, "Seeking solutions in configurable computing", IEEE Computer, December, 1997, pp. 38-43.
3. Keith Regester, Jong-Ho Byun, Arindam Mukherjee, and Arun Ravindran, Implementing Bioinformatics Algorithms on Nallatech-Configurable Multi-FPGA Systems, Xcell Journal, First Quarter 2005
4. Alexander Genkin, Casimir A. Kulikowski, Ilya Muchnik Set covering submodular maximization: An optimal algorithm for data mining in bioinformatics and medical informatics Journal of Intelligent and Fuzzy Systems, Special Issue: Challenges for future intelligent systems in

biomedicine, Volume 12, Number 1 / 2002, Pages: 5 - 17
5. Jie Zheng, Timothy J. Close, Tao Jiang and Stefano Lonardi, Efficient selection of unique and popular oligos for large EST databases Bioinformatics Vol. 20 no. 13 2004, pages 2101-2112
6. M.R. Garey, and D.S. Johnson, Computers and Intractability: A Guide tothe Theory of NPCompleteness, New York, NY: W.H.Freeman, 1979.
7. D.S. Johnson, "Approximation Algorithms for Combinatorial Problems", Journal of Computer and System Sciences, 1974, pp. 256-278.
8. S. Rampone, "Probability-driven Greedy Algorithms for Set Cover", in Proc. VIII SIGEF Congress "New Logics for the New Economy", Naples, Italy, September, 2001.
9. S. Dhar, M.Q. Rieck, S. Pai, and E.J. Kim, "Various Distributed Shortest Path Routing Strategies for Wireless Ad Hoc Networks", in Proc. 5th Int. Work. on Distributed Computing - Lecture Notes in Computer Science, 2918, Springer Verlag, 2003.
10. S. Dhar, M.Q. Rieck, S. Pai and E.J. Kim, "Distributed Routing Schemes for Ad Hoc Networks Using d-SPR Sets", Journal of Microprocessors and Microsystems, Special Issue on Resource Management in Wireless and Ad Hoc Mobile Networks, vol. 28(8), 2004, pp. 427-437.