IRis

Northeastern University

Institute for Complex Scientific Software Reports

Institute for Complex Scientific Software

January 01, 2005

End of year report : FY 2005

Gene Cooperman
Northeastern University

Institute for Complex Scientific Software

Recommended Citation

Cooperman, Gene and Institute for Complex Scientific Software, "End of year report : FY 2005" (2005). *Institute for Complex Scientific Software Reports*. Paper 2. http://hdl.handle.net/2047/d10008007

This work is available open access, hosted by Northeastern University.

INSTITUTE FOR COMPLEX SCIENTIFIC SOFTWARE

Gene Cooperman

End of Year Report FY 2005

1 GRANTS

1. **David Kaeli and Gene Cooperman**, National Science Foundation grant number ACIR-0342555, "Collaborative Research: Tuning Libraries to Effectively Exploit the Memory Hierarchy", (joint with Misha Kilmer (Tufts U.)), 2004-2007, $ 201,646.00 (Northeastern U. portion only)

2. **David Lorenz (CCIS), Paul Attie (CCIS), and Dana Brooks (ECE)**, "SoD: Design Locality: A Concept for Controlling the Design Complexity of Large Software Systems,", National Science Foundation, Science of Design Program (SoD), NSF Award 0438971, Dec. 2004 – Nov. 2007, $450,000.

2 GRANT PROPOSALS

The following are either pending, or were unsuccessful.

1. Lorenz (CCIS), Brooks (ECE), Kaeli (ECE), Cooperman (CCIS), (to NSF, Science of Design);

2. Kaeli (ECE) et al. (to NIH)

3. Ondrechen (Chemistry) and Williams (CCIS) (to NIH, "Functional Site Prediction in Protein Structure")

3 ACTIVITIES

1. **TITLE:** Computer Cluster

   PI: Computer Cluster, (supervised by Waleed Meleis (ECE))

   Activities: Over the past year, ICSS provided support for two parallel computing clusters. These clusters are an important computational resource for the center, and they are heavily used to perform computationally intensive simulations as well as fundamental computing research.

   The two parallel compute clusters were supported by ECE undergraduate students Kyle Guilbert and Matt Kowalski. Kyle and Matt provided a wide range of computer support services to the center, primarily on the joulian and keys compute clusters. Their main tasks included ensuring the availability of the clusters, installing and updating software, performing security and software audits, adding and maintaining user accounts, and evaluating the cost of gigabit networking equipment. Over the summer, we plan to perform software and hardware upgrades on the joulian cluster, including installing a new switch to replace the one that no longer works.

2. **TITLE:** Computer Architectural Barriers to High Performance Computation

   PI: Gene Cooperman (CCIS) and David Kaeli (ECE)

Supported: Xiaoqin Ma (5 hrs/wk) and Ye Wang (10 hrs/wk) (PhD Students), Summer I, 2005; Yijian Wang (20 hrs/wk) (Phd Student, ECE)

Activities: Cooperman and Kaeli are leading a research working group concerning issues of computer architectural barriers posed by RAM, disk and network. Some interests include compiler transformations and the use of our MBRAM model to predict running times of memory-bound programs. This support enabled Diego Rivera to begin a literature survey of the previous work in Sparse Matrix representations.

They also analyze the interactions of CPU, RAM, local disk, and network in a cluster. This cluster is becoming the standard high performance computer at the departmental level. Any improvements will have wide impact. This research is a follow-on to the successful NSF grant proposal of Cooperman and Kaeli through ICSS.

3. **TITLE:** Constructing Large Scientific Software using Pairwise Composition of Software Components

PIs: Robert Futrelle (CCIS)

Supported: Summer, '04: Rachel Utain-Evans and Carissa Kemp (undergraduates) and Mingyan Shao (PhD student); Fall/Spring, 2004–05: Mischa Bauermeister (undergraduate)

Activities: The support was critical to the success of a number of important projects, as well as contributing to papers and proposals. Rachel Utain-Evans did an excellent job of tracking down thousands of authors and email addresses to help launch the new BioMed Central journal, Biological Knowledge. This is non-commercial. As Editor-in-Chief, I receive no remuneration. Mingyan Shao made excellent progress on her dissertation research, contributing to diagram system software. Resulted in one paper accepted, and the groundwork for my 5/2005 NSF proposal. Another paper is under consideration. Carissa Kemp, now at Harvard, worked with me on the analysis of Biology papers' discourse structure. One paper rejected, currently under revision for another journal submission. This work will support my 7/2005 NSF and 10/2005 NIH proposals.

Mischa Bauermeister worked with a graduate student on getting Andrea Grimes' Framework Viewer running under Eclipse. This project had modest success. Development has ended, but it is still a useful system. Tim Hawes is a Linguistics senior who is absolutely superb. He single-handedly built a brand new totally object-oriented infrastructure for dealing with complex Biology text gotten from BioMed Central in XML format. His work will form the foundation of all our future NLP research, papers, proposals, etc.

4. **TITLE:** Accelerating Scientific Software with FPGA Hardware

PI: Miriam Leeser (ECE)

Activities: This is joint work with Dana Brooks (ECE), Charles DiMarzio (ECE) and Carol Warner (Biology)

Activities: We are developing reconfigurable hardware to accelerate phase unwrapping used for quadrature tomography microscopy (QTM) by Prof. Charles DiMarzio. QTM is used to image mouse embryos to further the research of Prof. Carol Warner. This is part of a larger effort using the Keck microscope to acquire data.

Phase unwrapping is currently done in software on a PC, but runs very slowly. Seven different algorithms are used for phase unwrapping, and in most cases all seven algorithms are run and the best results are selected. The FPGA implementation implements the residue calculation which is common to all seven algorithms. It is expected to significantly speedup phase unwrap.

The student working on this research plans to complete the implementation during the summer. For personal reasons, she was unable to do so during the spring semester, and will conclude this project with no further funds.

We expect to have measured speedup, an integrated system that works with the Keck microscope, and a submitted paper on this research by the end of August, 2005.

Supported: Heidi Mecklenburg (MS student)

5. **TITLE:** Constructing Large Scientific Software using Pairwise Composition of Software Components

   PIs: David Lorenz (CCIS) and Paul Attie (CCIS)

   Supported: Serguei Kojarski

   Activities: This work was also joint with Brooks, Cooperman and Kaeli, who provided application testbeds. The project addresses the problem of devising mechanical methods for synthesis and verification of large component-based scientific software. The major impediment to such mechanical analysis is the state-explosion problem. To avoid state-explosion we represent a system comprising of a set of components C_1, C_2, ... C_n in a pairwise normal form: the code that synchronizes a pair C_i, C_j of components is cleanly separated from the code that synchronizes other (even overlapping) components. This project resulted in a successful $450,000 NSF grant in 2004-2005.

6. **TITLE:** Development of Functional Genomics Methods with Machine Learning and THEMATICS

   PI: Mary Jo Ondrechen (Chemistry, A&S) and Ron Williams (CCIS))

   Activities: Our project deals with the integration of Machine methodologies and THEMATICS for powerful, new tools for the prediction of functional information for gene products. A smaller project on applications of our method to comparative model structures also included Valya Ilyin from the Biology Department.

   The purpose of this addendum is to provide more detail on the work of Wenxu Tong over this period since his name does not appear among the authors of the papers listed in the original progress report.

   The overall project deals with the automation of THEMATICS-based prediction of protein active sites, and the primary contribution that Wenxu Tong is making to this effort is the exploration of the application of machine learning methodologies using descriptors of the THEMATICS curves as input to trainable classifiers.

   Over the 2004-2005 time period, Wenxu Tong (PhD student, CCIS) has performed extensive studies of the use of Support Vector Machines for this purpose. A journal paper detailing these results is now ready for submission, and a two-page abstract describing this work has been submitted to the Automated Function Prediction Special Interest Group meeting.

   Wenxu Tong is currently investigating the application of boosting, a quite different but also promising machine learning methodology, to the same problem. Part of this work involves augmenting the THEMATICS-based features with other spatial features that we hope will help improve the overall accuracy of the classification process over that obtained using THEMATICS-based features alone.

   Supported: Terry Yang (PhD student, A&S); Wenxu Tong (PhD student, CCIS); and Heathew Brodkin (PhD student, A&S).

4 PUBLICATIONS ACKNOWLEDGING ICSS SUPPORT

(in alphabetic order)

1. D. Kaeli and Y. Wang, "Profile-guided File Partitioning on Beowulf Clusters", *Journal of Cluster Computing*, Special Issue on Parallel I/O, 2004.

2. D. Kaeli and Y. Wang, "Execution-Driven Simulation of Network Storage Systems," *Proceedings of the 12th ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS)*, October 2004, pp. 604–611.

3. D. Kaeli, W. Meleis, and J. Zhang,"Digital tomosynthesis mammography using a parallel maximum likelihood reconstruction method", Medical Imaging, San Diego CA, 2004.

4. J. Ko, P. Andre, L.F. Murga and M.J. Ondrechen, "Statistical Metrics for Protein Active Site Prediction with THEMATICS", Proceedings of the Fourth International Conference on Bioinformatics of Genome Regulation and Structure BGRS 2004, 1, pp. 282–285, Novosibirsk, Russia (2004).

5. J. Ko, L.F. Murga, P. Andre, H. Yang, M.J. Ondrechen, R.J. Williams, A. Agunwamba, and D.E. Budil, "Statistical Metrics for Protein Active Site Prediction with THEMATICS", *Proteins: Structure Function Bioinformatics* **59**, pp. 183–195 (2005).

6. J. Ko, L.F. Murga, Y.Wei, and M.J. Ondrechen, "Prediction of Active Sites for Protein Structures from Computed Chemical Properties", *Bioinformatics*, accepted.

7. D.H. Lorenz and S. Kojarski, "Pluggable AOP: Designing Aspect Mechanisms for Third-party Composition". *Proc. of Object-Oriented Programming Systems, Languages and Applications (OOPSLA)*, 2005. To appear.

8. L.F. Murga, J. Ko, Y. Wei, and M.J. Ondrechen, "Central Moments Based Statistical Analysis for the Determination of Functional Sites in Proteins with THEMATICS", submitted.

9. L.F. Murga, Y. Wei, P. Andre, J.G. Clifton, D. Ringe and M.J. Ondrechen, "Physicochemical Methods for Prediction of Functional Information for Proteins", Israel Journal of Chemistry **44**, 299–308 (2004).

10. D. Ng, D. Kaeli and D. Lorenz, "Program Comprehesion Using Aspects," Proceedings of the Workshop in Directions in Software Engineering Environments, held at ICSE 2004.

11. M.J. Ondrechen, "Identifying Functional Sites Based on Prediction of Charge Group Behavior," Protocols in Bioinformatics, A.D. Baxevanis, et al., Editors. 2004, John Wiley & Sons: Hoboken, N.J. p. 8.6.1 – 8.6.10.

12. M. Shao, D.E. Crispell, and R.P. Futrelle, "Moment-Derived Object Models for Vectorization", to appear in *Machine Vision Applications*, 2005.

13. I.A. Shehadi, A. Abyzov, A. Uzun, Y. Wei, L.F. Murga, V. Ilyin, and M.J. Ondrechen, "Active Site Prediction for Comparative Model Structures with THEMATICS," Journal of Bioinformatics and Computational Biology **3**, 127–143 (2005).