

Nagiza F. Samatova, Ph.D.

Associate Professor

Department of Computer Science

North Carolina State University

Senior Research Scientist

Computer Science and Mathematics Division

Oak Ridge National Laboratory

Voice: (919) 513-7575, Fax: (919) 515-7925

E-mail: samatovan@ornl.gov

Date: August 15, 2008

Contents

I. Introduction	1
I.A Brief Resume	1
I.B Self-Assessment	4
II. Teaching and Mentoring of Undergraduate and Graduate Students	8
II.A Teaching Effectiveness	8
II.B Instructional Development	10
II.D Master's and Doctoral Theses Directed and Being Directed	12
II.E Academic Advising	14
II.F Mentoring Activities	16
III. Scholarship in the Realms of Faculty Responsibility	18
III.A Scholarly Accomplishments	18
III.B Research Project Record	40
III.C Participation in Centers, Consortia and Other Organized Scholarly Efforts	43
IV. Extension and Engagement with Constituencies Outside the University	45
V. Technological and Managerial Innovation	48
VI. Service to the University and Professional Societies	49

I. Introduction

A. Brief Resume

1. Education

COMPUTATIONAL CENTER OF RUSSIAN ACADEMY OF SCIENCES, MOSCOW, RUSSIA, PhD, Applied Mathematics, 1993

UNIVERSITY OF TENNESSEE, KNOXVILLE, MS, Computer Science, 1998

TASHKENT STATE UNIVERSITY, UZBEKISTAN, BS, Applied Mathematics, 1991

SMITH COLLEGE, NEW ENGLAND, Certificate of Molecular Biologist, 1995

2. Professional Experience

NORTH CAROLINA STATE UNIVERSITY, Department of Computer Science, 2007-present
Associate Professor, 2007-present

OAK RIDGE NATIONAL LABORATORY, Computer Science and Mathematics Division
Senior Research Scientist, Joint Faculty Appointee, 2007-present
Senior Research Scientist, 2005-2007
Research Scientist, 2000-2004

UNIVERSITY OF TENNESSEE, KNOXVILLE
Adjunct Faculty, UT-ORNL Graduate School of Genome Science and Technology, 2003-present
Research Associate, Joint Institute for Computational Sciences (JICS), 1999-1999
Graduate Research Associate, Joint Institute for Computational Sciences (JICS), 1997-1999
Graduate Teaching Associate, Department of Mathematics, 1996-1997

OAK RIDGE ASSOCIATED UNIVERSITIES (ORAU), Computer Science and Mathematics Division
Postdoctoral Research Scientist, 1999-2000

TASHKENT STATE UNIVERSITY, UZBEKISTAN, Department of Applied Mathematics and Mechanics
Associate Professor, 1995-1996
Assistant Professor, 1993-1995

3. Scholarly and Creative Activities

Publications	Career Total	Total Over 7/1/07-6/30/08	Total since joining NCSU	Submitted [†]
Book and Book Chapter	3	1	1	
Refereed Journal Article	26	4	4	13
Refereed Conference Article	40	6	6	3
Invited Conference Article	5	3	3	
Invited Workshop Report	16	3	3	
Featured Article	3	2	2	
Patent	1	1	1	
Tutorial	2			
Technical Report	6	1	1	

[†]Not included in Career Total, Total Over 7/1/07-6/30/08 Period, or Total since joining NCSU Period

Presentations	Career Total	Total Over 7/1/07-6/30/08	Total since joining NCSU
Panels	2	2	2
Invited/Contributed	140	17	17

Funding	Career Total	Total Over 7/1/07-6/30/08	Total since joining NCSU	Pending
Grants (NCSU)	\$2,628,000	\$2,628,000	\$2,628,000	\$450,000
Grants (All)	\$28,484,000	\$3,930,000	\$3,930,000	\$1,500,000

Mentoring	Career Total	Total Over 7/1/07-6/30/08	Total since joining NCSU	In Progress[†]
Research Scientist	4	2	2	
Postdoc	4	1	1	
Postmaster	2			
IT Specialist	5			
PhD chair/co-chair	1			6
MS chair/co-chair				2
Written qualifier chair/advisor	4	4	4	1
Oral qualifier chair/advisor	1	1	1	3
PhD external committee	6	1	1	
PhD committee	2	2	2	2
MS committee	2	2	2	
High School Math Thesis	7	4	4	
Summer intern	23	7	7	1

[†]Not included in Career Total, Total Over 7/1/07-6/30/08, or Total since joining NCSU Period

Courses Taught	Career Total	Total Over 7/1/07-6/30/08	Total since joining NCSU
Regular undergraduate	4		
Discrete mathematics	2		
Mathematical logic	1		
Linear programming	1		
Regular graduate	3		
CS 594/UTK	1		
Graph theory	1		
Coding theory	1		

Courses Created/Revised in a Significant Way	Career Total	Total Over 7/1/07-6/30/08	Total since joining NCSU
Undergraduate	1		
Discrete mathematics	1		
Graduate	6	1	1
CSC 707 (created)	1	1	1
LSD 510/UTK (created)	1		
LSD 695 (created)	1		
Advanced graph theory (created)	1		
Advanced discrete mathematics(created)	1		
Special topics in 3-polytopes(created)	1		
Tutorials	2		

4. Membership in Professional Organizations

MEMBER, ACM Computing

5. Scholarly and Professional Honors

SIGNIFICANT EVENT AWARD, *Sustained Excellence in Mentoring that Leads to Multiple National Students Awards*, Battelle, 2007 *New in 2008*

2007 NOMINATION FOR UT-BATTELLE AWARDS NIGHT, EXCEPTIONAL COMMUNITY OUTREACH AWARD, Battelle, 2007 *New in 2008*

2007 YWCA TRIBUTE TO WOMEN FINALIST IN EDUCATION AND ADMINISTRATIVE PROFESSIONAL CATEGORY, Knoxville, 2007 *New in 2008*

OUTSTANDING MENTOR AWARD, UT-Battelle, 2008 *New in 2008*

INCENTIVIZED PERFORMANCE AWARD (IPA), UT-Battelle, 2006

2002 NOMINATION FOR UT-BATTELLE AWARDS NIGHT, SCIENTIFIC ACCOMPLISHMENT, EARLY CAREER, UT-Battelle, 2002

DISTINGUISHED PERFORMANCE EVALUATION, UT-Battelle, 2001-2007

SOFTWARE AGENT RESEARCH AWARD, *Virtual Information Processing Agent Research (VIPAR) Project, U.S. Pacific Command (USPACOM)*, Lockheed-Martin, 2001

APPRECIATION FOR EXCEPTIONAL WORK ON VIPAR PROJECT, Advanced Computing Technologies, ORNL, 2001.

6. Professional Service On Campus

Computing Infrastructure Committee, 2007-present *New in 2008*

Planning for Increasing Research Funds Committee, 2007-present *New in 2008*

7. Professional Service Off Campus (Most Recent)

Co-Chair, DOE/NSF Mathematics for Analysis of Petascale Data, June 2008 *New in 2008*

Co-Chair, DOE Genomics:GTL Program Systems Biology Network/Knowledgebase, May 2008 *New in 2008*

Co-Chair, OE/OASCR Mathematics for Peta-scale Data Analysis & Computation, June 2007 *New in 2008*

Technical Session Chair, DOE Workshop on Modeling and Simulation at the Exascale for Energy and the Environment, Apr, May, June 2007 *New in 2008*

Chair, DOE SciDAC Biology Summit, August 2005

co-Chair, DHS Threat Awareness Customer Data Integration & Dissemination, November 2005

co-Chair, DHS Data Sciences for Homeland Security Information Management and Knowledge Discovery, September 2004

MEMBER, various conference programs committees

DOE/DHS/NSF PANELIST on many occasions

REVIEWER for books, journals, and conferences

B. Self-Assessment

Research

Scientific computing (or computational science) is my core research focus area. Largely driven by the Department of Energy’s bioenergy and bioremediation mission, my fundamental and applied research linked foundational graph-theoretical modeling, statistical and high-performance computing, and computational biology. The inherently interdisciplinary nature of my research has not only advances the state-of-the-art in individual disciplines but ensures that technologies derived from this research impact real-world problems. A brief summary of some of these contributions follows next, with references to publications in Section III.A.

Graph Theory and Parallel Graph Algorithms. Many real-world problems, including those in biology can be modeled as graph problems. Yet due to their computational intractability (e.g., NP -hard nature), the applicability of graph algorithms is often limited to relatively small problems. Jointly with my research team, which I have assembled, I have addressed this challenge through fundamental theoretical studies and the development of scalable parallel graph algorithms, which significantly increase the range of their applicability.

I have introduced *graph perturbation theory* to solve graph problems on the “perturbed” graphs given a solution to the “baseline” or “unperturbed” graph [90,37-40,3-4]. ***Our graph perturbation theory has established a foundation for providing efficient (sometimes an order of magnitude faster) solutions for “perturbed” graphs*** by re-using the solution known for its “baseline” graph and finding the difference from the “baseline” solution rather than completely recomputing a new solution. Such performance improvement techniques are particularly useful for biological graphs derived from *uncertain* and *noisy* data, which will likely stimulate active research in the community.

To further scale algorithms up to larger graphs, I have led the development of parallel algorithms [5,30,44-45,48-49,57,61-62,87,90-91]. Parallelization has been non-trivial due to a number of challenges inherent to the nature of the combinatorial search space in such algorithms. On the one hand, graph algorithms often create a backtracking search tree with a highly irregular and hard to predict structure; therefore, almost any static decomposition of the search tree by parallel processors results in highly unbalanced processor execution times. On the other hand, the data-intensive nature of the search space, especially for enumeration problems, makes naive dynamic load distribution strategies that require extensive data movement prohibitively expensive. Specifically, maximal clique enumeration (MCE) is non-trivial to parallelize. It is ubiquitous in many applications, especially those in biology, yet these applications require the algorithms that can handle graphs with thousands or even millions of nodes. To the best of my knowledge, ***our parallel MCE algorithm is the first parallel implementation of the MCE algorithm; it also scales linearly to thousands of processors on distributed memory multi-core supercomputers***, such as Cray XT systems [5,48,]. The proposed MCE performance improvement techniques have general applicability to other graph problems that rely on backtracking search.

For certain classes of graphs (namely, 3-connected planar graphs), I have (a) *constructively* proven equal low and upper bounds on the maximum size of the vertex cover [35-36] and the face cover [41] in the class [39]; (b) provided a *constructive* proof for the *long-standing open problem* on the number of metric parameters that are necessary and sufficient for defining a polytope in 3-dimensional space given its graph [39,42], and (c) established the *asymptotic* upper bound on the number of equivalence classes of isomorphic graphs generated by arbitrary infinitesimal perturbations of the normal vectors to hyperplanes defining a convex polyhedron with its “baseline” graph provided [37-39].

High-Performance Statistical Computing. Scientific data sets are often massive, high-dimensional, noisy, geographically distributed, and/or dynamically generated. For distributed data, I have developed a hierarchical clustering method, named *RACHET* [32,82]. ***RACHET runs in at most linear $O(n)$ time,***

space, and communication costs to build a global hierarchy comparable in quality to centralized clustering by merging locally generated clustering hierarchies for n data points in d -dimensional space. This contrasts with centralized clustering algorithms that require bringing all the data into a centralized warehouse at $O(nd)$ transmission cost, which is prohibitively expensive for large data sets. *RACHET*'s framework supports a wide class of centroid-based hierarchical clustering algorithms, such as centroid, medoid, and Ward.

Similar to *RACHET*, I co-led the development of effective distributed dimension reduction algorithms. Distributed Principal Component Analysis (PCA) performs local PCA on local data without any data movement between the distributed systems and then moves and merges the p local principal components from the s distributed sites into a global PCA [80]. Unlike centralized methods, *distributed PCA reduces data transfers from $O(nd)$ to $O(sp)$, while controlling a desirable level of accuracy*. I have led the design of *a linear $O(n)$ time, accurate and robust dimension reduction algorithm, called RobustMap, with minimal data transmission* [76]. *RobustMap* selects a sequence of $p \leq d$ orthogonal axes defined by distant pairs of points, or pivots, in d -dimensional space and projects the data points into p -dimensional space. With my collaborators, I have proved that the pivots are the vertices of the convex hull of the data points in the original space. Therefore, *RobustMap provides robust means for multivariate outlier detection* [29,68].

In addition to founding distributed statistical computing methods, my research team has developed methods for analyzing continuous streams of data. *RobustMap* was generalized into the *Xmap algorithm to efficiently maintain a well-approximated, low-dimensional map of high-dimensional data streams at any given time* [75]. We have developed methods for maintaining a fixed-size random sample from streaming data, including *reservoir sampling with replacement that have sub-linear time complexity* [19,67].

While highly valuable, approaches that optimize the efficiency of individual algorithms are obviously low-throughput; they are tedious to design, implement and validate. *I proposed a complementary strategy through the development of parallel R (pR) for high performance statistical computing* [60,83,92-93]. *R* is an open source and widely-used environment for statistical computing. Its high-level data analytics language offers ease-of-use, but its scripting nature limits its scalability to massive data sets. *pR* is a middleware between *R* and a generic, parallel, optimized, and portable computational analytics engine. *pR* explores both data parallelism [60] and task parallelism (addressed by my collaborators in [13,112,50]) in *R*. Through *pR* a user can set up a parallel environment, distribute data and carry out the required parallel computation while maintaining the same look-and-feel interface of the *R* system. Its scalability is defined by the scalability of the underlying third-party parallel algorithms, such as *ScaLAPACK*, with minimal overhead introduced by the parallel *R* middleware. *The RScalAPACK package [60] is distributed across nearly 30 mirror sites around the world*.

Bioinformatics and Computational Biology. The theories and algorithms described above have been largely driven by a number of computational biology challenges that I have been addressing in collaboration with domain scientists. High-throughput mass spectrometry proteomics data is inherently noisy and presents a challenge to identify and quantify the relative abundances of thousands of proteins between two cellular conditions. To address this challenge, I led a research team to develop a computer program *ProRata* to automate the entire data analysis process for both stable-isotope-labeled [23-24,55] and label-free [22] quantitative shotgun proteomics. *ProRata outperformed previous quantitative methods, in addition to being the first program that provided quantitative confidence estimation (i.e., an "error bar") for each protein abundance ratio*. *ProRata* improved reconstruction of metabolic pathways in the hydrogen producing bacteria, *Rhodospseudomonas palustris* [17,56], and in the acid mine drainage microbial community [1]. *ProRata has been downloaded > 1000 times since 06/2006*.

My contributions have advanced the field of computational systems biology in a number of ways. They include, but are not limited to, (a) the study of functional consequences of stochastic fluctuations in molecular populations for gene regulation [25,51] and efficient algorithms for exact stochastic simulation of biochemical systems [26,52,63-64]; (b) an algorithm to discover substrate-specificity determining residues in a multi-functional protein enzyme family that are consistent with mutagenesis experiments [28,59,70] and its use for identifying CO₂/O₂ specificity-determining factors of the RuBisCo enzyme, which is important for carbon sequestration by bacteria [27]; and (c) algorithms for predicting protein-protein interactions and discovery of protein functional modules in protein interaction networks [18,9-10,54,69,74,84].

As of August 15, 2008, these and other studies have been recorded in 74 peer-reviewed publications: 26 journal papers, 48 conference papers and a number of invited talks summarized in Section III.A. Many of these publications are in high impact media including *Nature*, *Analytical Chemistry*, *Molecular and Cellular Proteomics*, *Discrete Mathematics*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, etc. There are 13 journal and 3 conference papers currently under the peer-review and are available upon request.

Teaching and Mentoring

I feel the greatest satisfaction from helping educate, train and advise the students that work with me. Seeing my advisees succeed and flourish is one of my greatest rewards. Defending my PhD in Moscow at the age of 22, I became a strong believer in an early scientific career and have devoted my life to promoting high-quality science among high school seniors, undergraduates, graduates, and postdocs (Sections II.D-E). I have supervised 6 postdocs, 9 postmasters, 20 PhD, 2 MS, 6 undergraduate, and 9 high-school students. From a diversity perspectives, almost half are women; 5 students and 3 faculty are underrepresented minorities.

A high school team that I mentored placed 1st (\$109,000 award) as national team finalists in the 2006-07 Siemens Competition in Math, Science and Technology for their project entitled “Linking Supercomputing and Systems Biology for Efficient Bioethanol Production.” A previous high school team placed 4th (\$39,000 award) as national team finalists in the 2005-06 Siemens competition for their project entitled “Searching with Comprehension: Going beyond Google.” My students also recieved national acclaim for their semi-finalist award in the Intel Talent Search and the recognition by the USA Today’s 2007.

These accomplishments have been read into the Congressional Record. Senator Lamar Alexander commented on the floor of the Senate during his speech in support of the National Competitiveness Investment Act [Congressional Record, Senate, Page S11588, Dec. 8, 2006] commented that “Part of the reason these three students succeeded is they were able to connect with the work and expertise at the Oak Ridge National Laboratory [and their] lead adviser, Dr. Nagiza F. Samatova ...” Numerous acknowledgements in the news, such as *The New York Times*, and in press, such as *Science* magazine, have drawn attention to the importance of early science education and the role my teams have played (see Section IV.A).

During my first year at NC State (from August 2007), I offered graduate research assistantships to 6 PhD and 2 MS students. One student has passed his oral PhD qualifier exam and three students have passed their PhD written qualifier exams. Two MS students received job offers from Google and Microsoft, respectively.

Due to my joint NCSU-ORNL faculty appointment, I have only taught the advanced core graduate theory course on Automata, Languages and Computability Theory (CSC707). I developed (from scratch) the course materials and exercised non-traditional ways of teaching including various ‘brain-teasers,’ team competitions, the ‘ugliest proof’ and ‘speed-track problem solving’ competitions. I received high evaluations across all the evaluation criteria (Section II.A). The peer faculty evaluation by Prof. Savage, who has extensive work experience with NC State, has complemented to my class by saying that it was “***the best class I have ever attended as a peer reviewer***” (Sections II.A-C).

Funding

Capitalizing on the successes of my research team and scientific collaborators, I have been awarded and managed, as a PI or co-PI, \$28,484,000 in research funds since 2001, with \$1,360,000 subcontracted to NC State this year and \$1,328,000 additional funds committed to NC State over the next four years. The funding agencies sponsoring my research programs include DOE, DHS, NSF, and UT-Battelle (see Section III.B).

Professional Service

I was invited to serve on the Department of Energy Office of Advanced Scientific Computing Research (OASCR) Panel of distinguished members of the computational science community to identify computational science breakthroughs across the Office of Science's research programs. I have served on the DOE Joint ASCAC/BERAC Subcommittee on Modeling and Simulation for GTL; co-chaired several workshops sponsored by the DOE, DHS and NSF for developing roadmaps for future R&D activities in computational sciences, as detailed in Section VI.B; and served on a number of scientific conference program committees and DOE/DHS/NSF proposal review committees, as detailed in Sections III.C-D, IV.B, VI. My service contributions as an educator have been recognized by a number of awards (see Section III.A).

II. Teaching and Mentoring of Undergraduate and Graduate Students

A. Teaching Effectiveness

Since I came as a joint faculty between NCSU and Oak Ridge National Laboratory (ORNL) in the Fall 2007 semester, I have taught the following course:

1. CSC 707 – AUTOMATA, LANGUAGES AND COMPUTABILITY THEORY (GRADUATE)

At the end of each semester, NCSU Computer Science students rate the effectiveness of their teachers on a 1 – 5 scale, where 1 is poor and 5 is truly outstanding. My teaching evaluations are listed in the table below.

I have received high evaluations despite the fact that students perceive my courses as presenting significantly more difficulty than average. I have received a number of informal complements both from the students and the faculty that the students really enjoy the course, while learning quite a lot of new material. I enormously enjoy teaching this course and my exposure to such bright and motivated students in the class.

Table 1: Teaching Evaluations in the past three academic years[†]

COURSE	SEMESTER	CLASS SIZE	INSTRUCTOR EXPLAINED DIFFICULT MATERIAL WELL	INSTRUCTOR WAS EFFECTIVE TEACHER	COURSE IMPROVED MY KNOWLEDGE OF SUBJECT	OVERALL COURSE WAS EXCELLENT	OVERALL LABS WERE EFFECTIVE EXPERIENCE
CSC 707	SPRING 2008	14	4.36	4.57	4.86	4.5	N/A
DEPT AVG			3.86	4.02	4.22	3.96	N/A

[†]I only taught one course this year due to my joint faculty appointment obligations and the one-course teaching load requirement as a part of my start-up hiring package.

Anonymous Students' Feedback: CSC 707 Spring 2008

New in 2008

- I love this course!
- In a word, she is an excellent teacher.
- Well-organized and -delivered lectures; use of competitions was interesting and engaging; instructor would often review difficult topics if the class did not seem to be catching on. Homework grading was unusual.
- The informal atmosphere of the class was one of its strongest points- students were able to reinforce each others' understanding of the material, or discuss different approaches to a problem before learning the "right" one. This made it much more enjoyable.
- I really enjoyed your enthusiasm for this material. I don't think I will ever forget the image of you with your back to the board arbitrarily dividing strings for the pumping lemma. You were always engaged and excited about the class. Your bonus points kept everyone trying (harder than they might have), and provided an alternate environment to the traditional lecture. I liked them.
- Dr. Samatova was exceptionally dedicated to our learning. Her main focus during the semester was to make sure that we were learning the material. She also focused on practical learning of

skills to use the theory of computation in our future work environments. She was very responsive to forum posts, e-mails, and student office visits about core material. Dr. Samatova occasionally had minor errors in her slides, which she promptly fixed.

- This was an excellent course, especially considering that Dr. Samatova was teaching it for the first time.

Peer Evaluation: CSC 707 January 2008

New in 2007

Peer evaluation of a CSC 707 lecture, conducted on January 28, 2008 by **Prof. Carla Savage** and **Prof. Rada Chirkova**:

We (Carla Savage and Rada Chirkova) attended Nagiza Samatova's course CSC 707, "Automata, Languages and Computability Theory" on Monday, January 28, 2008 for the purpose of conducting a peer evaluation of her teaching.

The lecture was on finite automata and regular languages, including closure properties and the pumping lemma. It began with an informative review of the last lecture - by individual students (!) with two facts contributed by each student. Dr. Samatova knows the names of the students, which made the review go quickly and smoothly.

Samatova used the computer projector to show extremely well-conceived slides, designed to lead the students through difficult proofs with illuminating visuals and carefully chosen examples. She is teaching the course for the first time and has developed all of these materials herself. The slides are available to the students on the course web page.

There are carefully designed weekly homeworks. Since the instructor has no TA for the class, she is grading all of the assignments herself, in addition to preparing the lecture materials.

The lecture was great!

The students were lively and attentive throughout the class. Some comments and observations on how Dr. Samatova was able to achieve this:

- all students can see the names of the other students written on paper placards on the desks
- ask students questions during the lecture (by name)
- students seem relaxed and comfortable answering the instructor's questions, laughing at the instructor's jokes, attentive and nodding during the explanations of the lecture materials
- encourages questions from students during lecture
- very good examples on lecture slides, most of the examples are not from the textbook, nice writing of additional explanations on the whiteboard

A particularly innovative idea that the students enjoyed was to have fun competitions in class between student teams: a "speed track", 4-5 students per team, very nice setup, nice challenging questions to be solved quickly, good leads by the instructor to help the students succeed.

There is a well-designed web page for the course and the instructor is making use of all of the options available on Wolfware. Some specific comments and observations:

- online syllabus, including the lecture slides, good coverage of all the appropriate material
- clearly stated learning outcomes

- nice ideas for homework problems (e.g., submit the "ugliest" solution - as a way to make the students more discriminating in writing and reading proofs.)
- prompt and helpful responses to message-board questions
- encourages anonymous feedback from students

In summary, we found Dr. Samatova to be a dedicated and competent teacher, leading a stimulating, challenging, and enjoyable class.

One of us (Savage) would like to add: **this was the best class I have ever attended as a peer reviewer.**

B. Instructional Development

Course Development

I have developed a course in the key area of Theoretical Computer Science: Automata, Languages and Computability Theory. These course is an integral part of the Master and PhD of Computer Science degrees.

In this course, I strived to achieve a good balance between theory and practice, and I developed substantial problem solving components that provide students with effective hands-on experience.

1. CSC 707 – AUTOMATA, LANGUAGES AND COMPUTABILITY THEORY

New in 2007

This is an advanced graduate course on the theory of computation and complexity, that I am offering in the Spring 2007 semester. The course material will be adapted for the undergraduate level in the Fall 2008. There are 14 students currently enrolled in the graduate course in contrast to 4-6 students in previous enrollments. I anticipate the increase of the enrollment for this course in out-years.

CSC707 is theoretically quite a challenging course. Since the students have quite a heterogenous background in mathematics, especially in formal proofs, teaching the course requires a very special treatment of that issue to make the course effective across-the-board. I developed (from scratch) the course materials including lecture slides, lecture notes, practice exams, quizzes, solutions, discussions, etc. Particularly, I have exercised numerous non-traditional ways of teaching including various 'brain-teasers,' team competitions, the 'ugliest proof' and 'speed-track problem solving' competitions. In spite of such a highly interactive methodology, I managed to cover not only the typical set of topics required for this course but also to bring up more advanced topics on fixed parameter tractability, approximability, and real-world applications. I was particularly proud of the students' performance of a very high (close to 90 out of 100) class grade average on quizzes and exams, for which an average score of 60 or less has been reported by other instructors and/or universities. The difficulties of the exam materials tested were comparable.

C. Teaching Outside NCSU

Fall 1993, Discrete Mathematics (Undergraduate), Tashkent State University

Fall 1993, Special Topics in 3-polytopes (Graduate Seminar), Tashkent State University

Spring 1994, Mathematical Logic (Undergraduate), Tashkent State University

Spring 1994, Coding Theory (Graduate Seminar), Tashkent State University

Fall 1994, Discrete Mathematics (Undergraduate), Tashkent State University

Fall 1994, Graph Theory & Combinatorics (Graduate), Tashkent State University

Fall 1994, Special Topics in 3-polytopes (Graduate Seminar), Tashkent State University
Fall 1994, Discrete Math & its Applications (Graduate Seminar), Tashkent State University
Spring 1995, Coding Theory (Graduate Seminar), Tashkent State University
Spring 1995, Linear Programming/Optimization (Undergraduate), Tashkent State University
Fall 1995, Discrete Mathematics (Undergraduate), Tashkent State University
Fall 1995, Graph Theory & Combinatorics (Graduate), Tashkent State University
Fall 2002, LSD 695: Advanced Computat. Methods in Bioinformatics (Graduate GST), U. of Tennessee (with Dr. Y. Xu)
Fall 2003, CS 594: Algorithmic Methods for Bioinformatics (Graduate CS), U. of Tennessee (with Prof. M. Langston)
Spring 2004, LSD 510: Bioinformatics (Graduate GST), U. of Tennessee (with Prof. Guo)

D. Master's and Doctoral Theses Directed and Being Directed

Doctoral Theses in Progress

1. AWEKAR, AMIT (PhD), NCSU
Topic: Mining Collaborative Knowledge Repositories
Sponsorship: Teaching Assistant for CSC333 *New in 2008*
2. BREIMYER, PAUL (PhD), NCSU
Topic: A Transparent Collaborative Framework for Efficient Data Analysis and Knowledge Annotation on the Web
Sponsorship: Graduate Research Assistant for a Research Grant *New in 2008*
3. CHEN, WEN BIN (PhD), NCSU
Topic: The Algorithms and Complexity of Multi-MCS
Sponsorship: Graduate Research Assistant for a Research Grant *New in 2008*
4. GREEN, NATHAN (PhD), NCSU
Topic: WebBANC: Building Semantically-Rich Annotated Corpora from Web User Annotations of Minority Languages
Sponsorship: Graduate Research Assistant for a Research Grant *New in 2008*
5. HENDRIX, WILLIAM (PhD), NCSU
Topic: Methodology for Database-Assisted Graph-Theoretical Algorithms for Perturbed Graphs
Sponsorship: Graduate Research Assistant for a Research Grant *New in 2008*
6. SCHMIDT, MATTHEW (PhD), NCSU
Topic: Computational Methods for Comparative Analysis of Large-Scale Networks
Sponsorship: Graduate Research Assistant for a Research Grant *New in 2008*

Master's Theses in Progress

1. KUMAR, VINAY (MS)
Topic: Scientific Social Networks for Collaborative Annotation
Sponsorship: Graduate Research Assistant for a Research Grant
2008 Google Summer Internship *New in 2008*
2. MOHAN, CHANDRA (MS)
Topic: In-database analytics with R
Sponsorship: Graduate Research Assistant for a Research Grant
Job offer from Microsoft *New in 2008*

Oral Preliminary Exam Committee Member

1. BREIMYER, PAUL, Defended on July 31, 2008
Title: A Transparent Collaborative Framework for Efficient Data Analysis and Knowledge Annotation on the Web
Advisor: Dr. Nagiza Samatova
Committee: Prof. Mladen Vouk, Dr. Tao Xie, and Dr. Steffen Heber *New in 2009*
2. LI, JIANGTIAN, Defended on May 6, 2008
Title: Run-time Automatic Parallelization for Statistical Computing
Advisor: Dr. Xiaosong Ma
Committee: Dr. Xiaohui (Helen) Gu, Dr. Frank Mueller, and Dr. Nagiza Samatova *New in 2008*

Written Preliminary Exam Committee Member

1. GREEN, NATHAN, Defended on August 8, 2008
Title: WebBANC: Building Semantically-Rich Annotated Corpora from Web User Annotations of Minority Languages
Advisor: Dr. Nagiza Samatova
Committee: Dr. James Lester (Area Representative) and Prof. Robert Rodman (Chair) *New in 2009*
2. HENDRIX, WILLIAM, Defended on April 18, 2008
Title: Theoretical Underpinnings for Maximal Clique Enumeration on Perturbed Graphs
Advisor: Dr. Nagiza Samatova
Committee: Prof. Carla Savage (Area Representative) and Prof. Jon Doyle (chair) *New in 2008*
3. SCHMIDT, MATTHEW, Defended on April 25, 2008
Title: Scalable, Parallel, and Distributed Memory Framework for Combinatorial Search and Enumeration in Modeling Large-Scale Biological Systems
Advisor: Dr. Nagiza Samatova
Committee: Dr. Donald Bitzer (Area Representative) and Dr. Edward Gehringer (Chair) *New in 2008*
4. ZHANG, ZHE, Defended on May 5, 2008
Title: Improving the Efficiency and Reliability of I/O Operations for High-End Computing Systems Storage Stacks.
Advisor: Dr. Xiaosong Ma
Committee: Prof. William Stewart (chair) and Dr. Nagiza Samatova (area representative) *New in 2008*

Master's Examination Committee Member

1. CHANDRA, HARSHA GIRISH, Defended on July 21, 2008
Title: Remote Data Collection and Analysis using Mobile Agents and Service-Oriented Architectg
Chair and Advisor: Dr. Frank Mueller
Committee: Dr. Helen Gu and Dr. Nagiza Samatova *New in 2009*
2. WHEELER, BENJAMIN, Defended on May 6, 2008
Title: Automating the annotation and discovery of Micro RNA in multispecies high throughput 454 Sequencing
Chair and Advisor: Dr. Steffen Heber
Committee: Dr. Nagiza Samatova and Dr. Brian Weigmann (Entomology Dept) *New in 2008*

Doctoral Dissertations Directed

1. PAN, CHONGLE, 2006
Dissertation: An Integrated Experimental and Computational Approach to Proteomics: Scaling from High Resolution Qualitative Analysis to Quantitative Measurements with Confidence Evaluation
Current Affiliation: ORNL, Research Scientist

E. Academic Advising

(1) Chair, PhD Written Qialifier Examination

1. COX, DAVID, November 2007

Advisor: Prof. Alan Tharp

Committee: Prof. Donald Bitzer (Area representative) and Dr. Nagiza Samatova (chair)

New in 2008

(2) PhD Dissertation Committees

1. LIN, HESHAN, in progress

Dissertation: High Performance Parallel and Distributed Genomic Sequence Search

2. LI, JIANTIAN, in progress

Dissertation: Run-time, Automatic Parallelization for Statistical Computing

(3) PhD Dissertation Co-advising Outside NCSU

1. ROCHA, ANDREA, open, USF

Dissertation: open

New in 2008

2. ZHANG, YUN, August 2008, UTK

Dissertation: Scalable Graph Algorithms for Biological Applications

Advisor: Prof. M. Langston, UTK

3. MCCOLLUM, MIKE, 2006, UTK

Dissertation: Parallel Exact Stochastic Simulation Algorithms

Advisor: Prof. G. Peterson, UTK

2004-2005 Mike McCollum (PhD, Electrical Engineering, UTK, TN. Advisor: Dr. Greg Peterson)

4. ABU-KHZAM, FAISAL, 2003, UTK

Dissertation: FPT Graph and Distributed Data Mining Algorithms

Advisor: Prof. M. Langston, UTK

(4) PhD Qualifying Exam Committees

1. RAZUMOVSKAYA, YEVGENIYA, October 2004, UTK *Topic:* Development of a framework for a computational method for identifications of post-translational modifications
2. XU, QIN, November 2004, UTK *Topic:* Understanding the catalytic mechanism of Rubisco on CO_2 and O_2 fixation and determination of the activation energies of these two processes

(5) High School Math and Sciences Theses Directed

1. ARCANGELLI, STEVEN, 2007, Oak Ridge High School

Thesis: Efficient parallel and scalable methods for systems biology

Current Affiliation: Massachusetts Institute of Technology (MIT)

2006-2007 National Finalist of the Siemens Math, Science, and Technology Competition, 1st place

2007 Southern Appalachian Science and Engineering Fair, 1st place

New in 2008

2. BRENT, PATRICIA, 2006, Oak Ridge High School

Thesis: Named Entity Recognition with Special Attention to Feature Selection and Convergence Criteria

Current Affiliation: University of Chicago

- 2005-2006 National Finalist of the Siemens Math, Science, and Technology Competition, 4th place**
3. GRABENSTEIN, NICHOLAS, 2006, Oak Ridge High School
Thesis: Going Beyond Google: Searching with Comprehension
Current Affiliation: Massachusetts Institute of Technology (MIT)
2005-2006 National Finalist of the Siemens Math, Science, and Technology Competition, 4th place
 4. GU, ALICE, 2007, Oak Ridge High School
Thesis: From Integrative Systems Biology to Improved Bioremediation Solutions
Current Affiliation: Washington University in St. Louis
2006-2007 Semi-finalist Intel Science Talent Search Competition
2007 Southern Appalachian Science and Engineering Fair, 2nd place *New in 2008*
 5. HORTON, SCOTT, 2007, Oak Ridge High School
Thesis: Computational Methods for Efficient Bioethanol Production
Current Affiliation: University of Virginia
2006-2007 National Finalist of the Siemens Math, Science, and Technology Competition, 1st place
2007 Southern Appalachian Science and Engineering Fair, 1st place *New in 2008*
 6. MOLONY, SCOTT, 2007, Oak Ridge High School
Thesis: Comparative Analysis of Biological Networks for Efficient Ethanol Production
Current Affiliation: Boston College, Full Scholarship
2006-2007 National Finalist of the Siemens Math, Science, and Technology Competition, 1st place
USA Today's 2007
2007 Southern Appalachian Science and Engineering Fair, 1st place *New in 2008*
 7. UMAR, TARIK, 2006, Oak Ridge High School
Thesis: Conflict Resolution for Named Entity Recognition
Current Affiliation: Harvard University
2005-2006 National Finalist of the Siemens Math, Science, and Technology Competition, 4th place

F. Mentoring Activities

Research Scientists

1. KARPINETS, TATIANA, 2006-2007
Current Affiliation: University of Tennessee, Joint Institute for Biological Science (JIBS)
2. KORA, GURUPRASAD, 2006-present
Current Affiliation: University of Tennessee, Joint Institute for Computational Science (JICS) *New in 2008*
3. PAN, CHONGLE, 2006-present
Current Affiliation: Oak Ridge National Laboratory, Research Scientist *New in 2008*
4. PARK, BYUN-HOON, 2005-present
Current Affiliation: Oak Ridge National Laboratory, Research Scientist

Postdoctoral Fellows

1. BALDWIN, NICOLE, 2004-2005
Topic: Integrated Knowledgebase for Shewanella Federation
Current Affiliation: University of Texas Health Science Center, Houston, TX
2. PARK, BYUN-HOON, 2002-2005
Topic: Advanced methods for protein-protein interactions
Current Affiliation: Oak Ridge National Laboratory
3. TIAN, WENHONG, 2007-2008
Topic: Allignment of multiple biological networks
Current Affiliation: North Carolina State University *New in 2008*
4. YU, GONG-XIN, 2003-2005
Topic: Multi-resolution functional characterization of protein families
Current Affiliation: Boise State University, ID, Assistant Professor
5. ZANG, BING, 2005-2006
Topic: Methods for label-free quantitative proteomics
Current Affiliation: Vanderbilt University, Assistant Professor

Postmaster Fellows

1. KRISHNAMURTHY, RAMYA, 2003-2007
Topic: Framework for Named Entity Recognition
2. SYMONS, CHRISTOPHER, 2004-2007
Topic: Active Learning Framework for Named Entity Recognition
Current Affiliation: Oak Ridge National Laboratory

IT Specialists

1. CHANDRAMOHAN, PRAVEEN, 2003-2005
2. BAUER, DAVID, 2001-2004
Current Affiliation: Georgia Tech, PhD student
3. KORA, GURUPRASAD, 2002-2006
Current Affiliation: University of Tennessee, Joint Institute for Computational Science (JICS)

4. MUNAVALLI, RAJESH, 2003-2005
5. WATKINS, IAN, 2001-2005

Summer Internships Directed

- | | |
|----------------------------------|--------------------|
| 1. JALLOUK, ANDREW, 2006-2007 | <i>New in 2008</i> |
| 2. HORTON, SCOTT, 2006-2007 | <i>New in 2008</i> |
| 3. ARCANGELLI, STEVEN, 2006-2007 | <i>New in 2008</i> |
| 4. KOTHE, BRETT, 2007 | <i>New in 2008</i> |
| 5. BRENT, PATRICIA, 2005-2007 | <i>New in 2008</i> |
| 6. GU, ALICE, 2006 | |
| 7. LI, JIANGTIAN, 2005-2006 | |
| 8. FRIDMAN, MAX, 2006 | |
| 9. UMAR, TARIK, 2005 | |
| 10. GRABENSTEIN, NICHOLAS, 2005 | |
| 11. OZISIKYILMAZ, BERKIN, 2005 | |
| 12. LIN, HESHAN, 2004 | |
| 13. BURTON, JEFFREY, 2004 | |
| 14. BRAIMAN, AVITAL, 2003 | |
| 15. BAUER, DAVID, 2001-2002 | |
| 16. HESPEN, JENNIFER, 2001-2002 | |
| 17. QU, YONGMING, 2001 | |
| 18. MIRONOVA, SVETLANA, 2001 | |

Black Colleges, Universities and Minority Educational Institutions Internships Directed

- | | |
|---------------------------------|--------------------|
| 1. ROCHA, ANDREA, 2007-2008 | <i>New in 2008</i> |
| 2. MCGARITY, JESSICA, 2006-2007 | <i>New in 2008</i> |
| 3. WALKER, KATHRYN, 2006 | |
| 4. ASKIA, RASHIDA, 2004 | |
| 5. LESLIE, TRAYVON, 2004 | |

III. Scholarship in the Realms of Faculty Responsibility

A. Scholarly Accomplishments

Awards and Honors

- SIGNIFICANT EVENT AWARD, *Sustained Excellence in Mentoring that Leads to Multiple National Students Awards*, Battelle, 2007 *New in 2008*
- 2007 NOMINATION FOR UT-BATTELLE AWARDS NIGHT, EXCEPTIONAL COMMUNITY OUT-REACH AWARD, Battelle, 2007 *New in 2008*
- 2007 YWCA TRIBUTE TO WOMEN FINALIST IN EDUCATION AND ADMINISTRATIVE PROFESSIONAL CATEGORY, Knoxville, 2007 *New in 2008*
- OUTSTANDING MENTOR AWARD, UT-Battelle, 2008 *New in 2008*
- INCENTIVIZED PERFORMANCE AWARD (IPA), UT-Battelle, 2006
- 2002 NOMINATION FOR UT-BATTELLE AWARDS NIGHT, SCIENTIFIC ACCOMPLISHMENT, EARLY CAREER, UT-Battelle, 2002
- DISTINGUISHED PERFORMANCE EVALUATION, UT-Battelle, 2001-2007
- SOFTWARE AGENT RESEARCH AWARD, *Virtual Information Processing Agent Research (VIPAR) Project, U.S. Pacific Command (USPACOM)*, Lockheed-Martin, 2001
- APPRECIATION FOR EXCEPTIONAL WORK ON VIPAR PROJECT, Advanced Computing Technologies, ORNL, 2001.
- THE FREEDOM SUPPORT ACT GRADUATE FELLOWSHIP, United States of America, 1996
- V. I. LENIN GRADUATE FELLOWSHIP, Tashkent State University, 1991-1993
- AL HOREZMI SCHOLARSHIP, Tashkent State University, 1988-1991
- 2nd PLACE WINNER, STATE (UZBEKISTAN) MATHEMATICS COMPETITION FOR HIGH SCHOOL STUDENTS, Tashkent, Uzbekistan, 1986
- 1st PLACE WINNER, STATE (UZBEKISTAN) CLASSIC GUITAR COMPETITION, Tashkent, Uzbekistan, 1984
- 1st PLACE WINNER, STATE (UZBEKISTAN) RHYME RECITATION COMPETITION, Tashkent, Uzbekistan, 1980

Publications Summary

The Chancellor has asked faculty to identify the extent of their contributions in their various publications. The following table summarizes this information. Each publication listed below is coded with one or more letters defined as follows:

C – Corresponding author

L – Lead author

E – Equal contribution as the other authors

M – Minor contribution as an author

Publications	Career Total	Corresponding [†] [C]	Lead [L]	Equal [E]	Minor [M]
Book and Book Chapters	3	2	1	2	
Refereed Journal Articles	26	18	10	14	2
Refereed Conference Articles	40	27	4	32	4
Invited Conference Articles	5	3	2	3	
Invited Featured Articles	3	2	2	1	
Invited Workshop Reports	16		2	9	5
Patents	1			1	
Tutorials	2	2	2		
Technical Reports	6	2	1	4	1
Submitted Journal Articles	13	10	1	9	2
Submitted Conference Articles	3	4		3	1
Total	118	70	25	78	15

[†]As a corresponding author, I define the problem, lead the research team to solve the problem, provide funding for these research studies, and contribute significantly to the manuscript writing and editing.

Submitted Journals Publications

- [M] BELNAP, C.P., PAN, C., VERBERKMOES, N.C., POWER, M.E., SAMATOVA, N.F., CARVER, R.L., HETTICH, R.L., BANFIELD, J.F., “Quantitative proteogenomic analysis of highly productive extremely acidophilic microbial communities.” Submitted to *Nature*. *New in 2009*
- [CE] PAN, C., HAYES, M.W., PARK, B.H., VERBERKMOES, N.C., BANFIELD, J.F., HETTICH, R.L., SAMATOVA, N.F., “Confident identification of over ten thousand de novo sequence tags from shotgun proteomic data using high-resolution Orbitrap tandem mass spectrometry.” Submitted to *J Proteome Research*. *New in 2009*
- [CE] HENDRIX, W., SCHMIDT, M., BREIMYER, P., SAMATOVA, N.F., “Theoretical underpinnings for maximal clique enumeration on perturbed graphs.” Submitted to *Theoretical Computer Science*. *New in 2009*
- [CE] HENDRIX, W., SCHMIDT, M., BREIMYER, P., SAMATOVA, N.F., “On perturbation theory and an algorithm for maximal clique enumeration in uncertain and noisy graphs.” Submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*. *New in 2009*
- [CE] SCHMIDT, M., THOMAS, K., PARK, B.-H., SAMATOVA, N.F., “A Scalable, parallel algorithm for maximal clique enumeration.” Submitted to *J. of Parallel and Distributed Computing*. *New in 2009*

6. [CE] CHEN, W., SCHMIDT, M., SAMATOVA, N.F., “On parameterized complexity of Multi-MCS problem.” Submitted to *Theoretical Computer Science*. New in 2009
7. [CE] CHEN, W., SCHMIDT, M., SAMATOVA, N.F., “The parameterized complexity of maximum common subtrees and minimum common supertrees problem.” Submitted to *Information Processing Letters*. New in 2009
8. [CE] KARPINETS, T.V., PELLETIER, D.A., PAN, C., UBERBACHER, E.C., MELNICHENKO, G.V., HETTICH, R.L., SAMATOVA, N.F., “Large-scale transcriptomics and proteomics reveal the involvement of syntrophic consortia of *Rhodospseudomonas palustris* in anaerobic degradation of aromatic compounds.” Submitted to the *BMC Genomics*. New in 2008
9. [CE] KARPINETS, T., PARK, B.H., ZHANG, B., PELLETIER, D.A., SCHMOYER, D., McDONALD, W.H., HURST, G., UBERBACHER, E., BUCHANAN, M.V., SAMATOVA, N.F., “Exploiting the blessing of noise in large-scale affinity isolation experiments for sensitive and specific identification of protein complexes.” Submitted to the *Proteome Science*. New in 2008
10. [CL] SAMATOVA, N.F., PARK, B.H., KARPINETS, T., “The common-target model reveal stress-related transcriptional reprogramming of yeast cells in pull-down experiments.” Submitted to the *Journal of Biological Engineering*. New in 2008
11. [CE] BREIMYER, P., SAMATOVA, N.F., KORA, G., “Web-Enabled R application requirements to support large-scale data-intensive collaborative statistical computing.” Under revision in *J. of Statistical Software*. New in 2008
12. [M] LIN, H., MA, X., FENG, W., SAMATOVA, N.F., “Scalable task and I/O scheduling for massively parallel genomic sequence searches.” Submitted to *Journal*. New in 2008
13. [CM] LI, J., MA, X., YOGINATH, S., KORA, G., SAMATOVA, N.F., “pR: Automatic, transparent runtime parallelization of the R scripting language.” Submitted to *IEEE TPDS*. New in 2008

Submitted Conference Publications

14. [CE] BREIMYER, P., GREEN, N., KUMAR, V., SAMATOVA, N.F., “BioDEAL: Biological data-evidence-annotation linkage system.” Submitted to *IEEE BIBM Conference*. New in 2009
15. [CE] GREEN, N., BREIMYER, P., KUMAR, V., SAMATOVA, N.F., “WebBANC: Building semantically-rich annotated corpora from Web user annotations of minority languages.” Submitted to *Conference on Empirical Methods in Natural Language Processing (EMNLP2008)*. New in 2009
16. [CE] BRENT, P., GREEN, N., BREIMYER, P., KRISHNAMURTHY, R., SAMATOVA, N.F., “Systematic evaluation of convergence criteria in iterative training for NLP.” Submitted to *Conference*. New in 2009

Published Journal Publications

17. [E] PAN, C., ODA, Y., LANKFORD, P.K., ZHANG, B., SAMATOVA, N.F., PELLETIER, D.A., HARWOOD, C.S., HETTICH, R.L., “Characterization of anaerobic catabolism of *p*-coumarate in *Rhodospseudomonas palustris* by integrating transcriptomics and quantitative proteomics.” *Mol Cell Proteomics*, vol. 7, no. 5, pp. 938-48, 2008, PMID: 18156135 (**Impact factor**: 9.62). New in 2008
18. [CE] ZHANG, B., PARK, B.H., KARPINETS, T., SAMATOVA, N.F., “From pull-down data to protein interaction networks and complexes with biological relevance.” *Bioinformatics*, vol. 24, no. 7, pp. 979-86, 2008, PMID: 18304937 (**Impact factor**: 4.9). New in 2008

19. [E] PARK BH, OSTROUCHOV G, SAMATOVA NF., "Sampling streaming data with replacement." *Comput. Stat. Data Anal.*, vol. 52, no. 2, pp. 750-762, Oct 2007, PMID: 18304937 (**Impact factor:** 0.93). New in 2008
20. [CE] SISNEROS, R., JONES, C., HUANG, J., GAO, J., PARK, B.H., SAMATOVA, N.F., "A multi-level cache model for run-time optimization of remote visualization." *IEEE Trans Vis Comput Graph*, vol. 13, no. 5, pp. 991-1003, Sep-Oct 2007, PMID: 17622682 (**Impact factor:** 1.8). New in 2008
21. [M] KARPINETS T, GREENWOOD DJ, POGRIBNY I, SAMATOVA N., "Bacterial stationary-state mutagenesis and Mammalian tumorigenesis as stress-induced cellular adaptations and the role of epigenetics." *Curr Genomics*, vol. 7, no. 8, pp. 481-96, Nov 2006, PMID: 18369407 (**Impact factor:** 0.79).
22. [CE] ZHANG B, VERBERKMOES NC, LANGSTON MA, UBERBACHER E, HETTICH RL, SAMATOVA NF., "Detecting differential and correlated protein expression in label-free shotgun proteomics." *J Proteome Res.*, vol. 5, no. 11, pp. 2909-18, Nov 2006, PMID: 17081042 (**Impact factor:** 5.2).
23. [CE] PAN C, KORA G, McDONALD WH, TABB DL, VERBERKMOES NC, HURST GB, PELLETIER DA, SAMATOVA NF, HETTICH RL., "ProRata: A quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation." *Anal Chem.*, vol. 78, no. 20, pp. 7121-31, Oct 2006, PMID: 17037911 (**Impact factor:** 5.45).
24. [CE] PAN C, KORA G, TABB DL, PELLETIER DA, McDONALD WH, HURST GB, HETTICH RL, SAMATOVA NF., "Robust estimation of peptide abundance ratios and rigorous scoring of their variability and bias in quantitative shotgun proteomics." *Anal Chem.*, vol. 78, no. 20, pp. 7110-20, Oct 2006, PMID: 17037910 (**Impact factor:** 5.45).
25. [E] AUSTIN DW, ALLEN MS, MCCOLLUM JM, DAR RD, WILGUS JR, SAYLER GS, SAMATOVA NF, COX CD, SIMPSON ML., "Gene network shaping of inherent noise spectra." *Nature*, vol. 439, no. 7076, pp. 608-11, Feb 2006, PMID: 16452980 (**Impact factor:** 26.7).
26. [CE] MCCOLLUM JM, PETERSON GD, COX CD, SIMPSON ML, SAMATOVA NF., "The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior." *Comput Biol Chem.*, vol. 30, no. 1, pp. 39-49, Feb 2006, PMID: 16321569 (**Impact factor:** 2.14).
27. [CE] YU GX, PARK BH, CHANDRAMOHAN P, GEIST A, SAMATOVA NF., "An evolution-based analysis scheme to identify CO₂/O₂ specificity-determining factors for Ribulose 1,5-bisphosphate carboxylase/oxygenase." *Protein Eng Des Sel.*, vol. 18, no. 22, pp. 589-96, Oct 2005, PMID: 16246824 (**Impact factor:** 3.0).
28. [CE] YU GX, PARK BH, CHANDRAMOHAN P, MUNAVALLI R, GEIST A, SAMATOVA NF., "In silico discovery of enzyme-substrate specificity-determining residue clusters." *J Mol Biol.*, vol. 352, no. 5, pp. 1105-17, Oct 2005, PMID: 16140329 (**Impact factor:** 4.9).
29. [E] OSTROUCHOV G, SAMATOVA NF., "On FastMap and the convex hull of multivariate data: toward fast and robust dimension reduction." *IEEE Trans Pattern Anal Mach Intell.*, vol. 27, no. 8, pp. 1340-3, Aug 2005, PMID: 16119272 (**Impact factor:** 4.3).
30. [CE] SUTERS WH, ABU-KHZAM FN, ZHANG Y, SYMONS CT, SAMATOVA NF, LANGSTON MA, "A new approach and faster exact methods for the maximum common subgraph problem." *Lecture Notes in Computer Science*, vol. 3595, pp. 717-727, 2005.
31. [E] HEFFELFINGER GS, MARTINO A, GORIN A, XU Y, RINTOUL MD 3RD, GEIST A, AL-HASHIMI HM, DAVIDSON GS, FAULON JL, FRINK LJ, HAALAND DM, HART WE, JAKOBSSON

- E, LANE T, LI M, LOCASCIO P, OLKEN F, OLMAN V, PALENIK B, PLIMPTON SJ, ROE DC, SAMATOVA NF, SHAH M, SHOSHANI A, STRAUSS CE, THOMAS EV, TIMLIN JA, XU D., "Carbon sequestration in *Synechococcus Sp.*: from molecular machines to hierarchical modeling." *OMICS*, vol. 6, no. 4, pp. 305-30, 2002, PMID: 12626091.
32. [CL] SAMATOVA NF, OSTROUCHOV G, GEIST A, MELECHKO AV., "RACHET: An efficient cover-based merging of clustering hierarchies from distributed datasets." *Distrib. Parallel Databases*, vol. 11, no. 2, pp. 157-180, Mar 2002, PMID: 16119272 (**Impact factor:** 1.03) (**Invited paper**).
33. [M] MELECHKO AV, SIMKIN MV, SAMATOVA NF, BRAUN J, PLUMMER EW , "Complex structural phase transition in a defect-populated two-dimensional system." *PHYSICAL REVIEW B*, vol. 64, no. 23, pp. 157-180, Dec 2001 (**Impact factor:** 3.1).
34. [CL] SAMATOVA NF, POTOK TE, LEUZE MR, "Vector space model for the generalized parts grouping problem." *Robot. Comput.-Integr. Manuf.*, vol. 17, no. 1-2, pp. 73-80, Feb-Apr 2001 (**Impact factor:** 0.81) (**Invited paper**).
35. [CL] SAMATOVA NF, "On the nonhomogeneity index of graphs of polytopes." *Diskr. Mat.*, vol. 6, no. 3, pp. 89-93, 1994, Moscow, Russia.
36. [CL] SAMATOVA NF, "On the nonhomogeneity index of graphs of polytopes." *Discrete Mathematics and Applications*, vol. 4, no. 4, pp. 371-375, 1994, Moscow, Russia.
37. [CL] SAMATOVA NF, PULATOV AK., "On the measure of combinatorial instability of the specification of a convex polyhedron in R^3 ." *Diskr. Mat.*, vol. 5, no. 3, pp. 150-156, 1993, Moscow, Russia.
38. [CL] SAMATOVA NF, PULATOV AK., "The estimation of the measure of combinatorial instability of a convex 3-polytope under limited deformations." *UzNIINTI* (Deposited), 1993, Tashkent, Uzbekistan, N 1798 - Uz93.
39. [CL] SAMATOVA NF, "On the complexity characteristics of the methods of specification and combinatorial irregularity of a convex 3-polyhedro." *Thesis for a Candidate's degree* (Computing Center of Russian Academy of Sciences, 1993, Moscow, Russia.
40. [CL] SAMATOVA NF, "On the measure of combinatorial unsteadiness of a 3-polytope under linear deformations." *UzNIINTI* (Deposited), 1993, Tashkent, Uzbekistan, N 1799 - Uz93.
41. [CL] SAMATOVA NF, SHIROKOVA OB, PULATOV AK., "Maximum values of some characteristics for vertex cover in a 3-polytope." *Methods of discrete analysis in graph theory and complexity*, vol. 52, pp. , 1992, Novosibirsk, Russia.
42. [L] PULATOV AK, SAMATOVA NF, "Complexity of the specification of a convex polyhedron in R^3 ." *Diskr. Mat.*, vol. 3, no. 2, pp. 148-156, 1991.

Published Refereed Conferences

43. [CE] PARK BH, ZHANG B, KARPINETS T, SAMATOVA NF, “Multi-stage Framework to Infer Protein Functional Modules from Mass Spectrometry Pull-Down Data with Assessment of Biological Relevance.” *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2007)*, pp. 223-229, ISBN: 978-0-7695-3031-4 doi: 10.1109/BIBM.2007.52, Nov 2-4, 2007. New in 2008
44. [CE] PARK B.H., SCHMIDT M., THOMAS T., SAMATOVA NF, “Parallel, Scalable, Memory-Efficient Backtracking for Combinatorial Modeling of Large-Scale Biological Systems.” *Proceedings of the 22nd IEEE International Parallel and Distributed Processing Symposium (IPDPS 2008), the 7th IEEE International Workshop on High Performance Computational Biology (HiCOMB 2008)*, pp. , Miami, Florida USA, Apr 14-18, 2008 New in 2008
45. [E] ABU-KHZAM, F.N., RIZK, M.A., ABDALLAH, D.S., SAMATOVA N.F., “The Buffered Work-Pool Approach for Search-Tree Based Optimization Algorithms.” *Proceedings of the 7th International Conference on Parallel Processing and Applied Mathematics (PPAM 2007)*, Gdansk, POLAND, September 9-12, 2007. New in 2008
46. [M] LIN H., MA X., LI J., YU T., SAMATOVA N.F., “Adaptive Request Scheduling for Parallel Scientific Web Services.” *Proceedings of the 20th International Conference on Scientific and Statistical Database Management (SSDBM '08)*, Hong Kong, Jul 9-11, 2008. New in 2008
47. [CE] PARK BH, SAMATOVA NF, MUNAVALLI R, KRISHNAMURTHY R, KETTANI H, GEIST A, “Rapid and Robust Ranking of Text Documents in a Dynamically Changing Corpus.” *Proceedings of the 6th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA-08), Databases and Data Mining Track*, pp. , Doha, Qatar, Mar 31 - Apr 4, 2008 New in 2008
48. [CE] THOMAS K, SAMATOVA NF, SCHMIDT M, PARK BH, “Efficient Scaling Up of Parallel Graph Algorithms for Genome-Scale Biological Problems on Cray XT.” *Proceedings of the Cray User Group (CUG-08) Conference, Crossing the Boundaries*, Helsinki, May 5-8, 2008. New in 2008
49. [CE] ABU-KHZAM FN, SAMATOVA NF, RIZK MA, LANGSTON MA., “The Maximum Common Subgraph Problem: Faster Solutions via Vertex Cover.” *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications (AICCSA07)*, pp. 367 - 373, 13-16 May 2007, Amman, Jordan, doi: 10.1109/AICCSA.2007.370907.
50. [CE] MA, X.; LI, J.; SAMATOVA, N.F., “Automatic Parallelization of Scripting Languages: Toward Transparent Desktop Parallel Computing.” *Proceedings of IEEE/ACS International Conference on Parallel and Distributed Processing Symposium (IPDPS 2007)*, pp. 1-6, 26-30 March 2007, doi: 10.1109/IPDPS.2007.370488.
51. [E] AUSTIN, D.; ALLEN, M.; MCCOLLUM, J.M.; DAR, R.D.; SAYLER, G.S.; SAMATOVA, N.F.; COX, C.D.; SIMPSON, M., “Gene network shaping of inherent noise spectra.” *Proceedings of the Bio Micro and Nanosystems Conference (BMN 2006)*, pp. 11-11, 15-18 Jan. 2006, doi: 10.1109/BMN.2006.330897.
52. [E] COX, C.D.; MCCOLLUM, J.M.; DAR, R.D.; AUSTIN, D.; ALLEN, M.S.; SAMATOVA, N.F.; SAYLER, G.S.; SIMPSON, M.L., “Calibration of a stochastic model of gene expression including feedback and extrinsic noise sources.” *Proceedings of the Bio Micro and Nanosystems Conference (BMN 2006)*, pp. 94-94, 15-18 Jan. 2006, doi: 10.1109/BMN.2006.330904.
53. [CE] SYMONS, C.T.; SAMATOVA, N.F.; KRISHNAMURTHY, R.; PARK, B.H.; TARIK UMAR; BUTTLER, D.; CRITCHLOW, T.; HYSOM, D., “Multi-Criterion Active Learning in Conditional

- Random Fields.” *Proceedings of 8th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2006)*, pp. 323-331, Nov 13-15, 2006, Washington D.C., doi: 10.1109/IC-TAI.2006.90.
54. [CE] PARK BH, KARPINETS T, ZHANG B, SAMATOVA NF, “Common-target model for identification of protein interaction modules in pull-down experiments.” *Proceedings of the Annual Meeting of the Institute for Biological Engineering*, Mar 30-April 1, St. Louis, Missouri.
 55. [CE] PAN C, KORA G, TABB D, McDONALD WH, PELLETIER D, HURST G, HETTICH R, SAMATOVA NF, “ProRata: a software package with improved point and interval estimation of protein abundance ratio for quantitative shotgun proteomics.” *Proceedings of the 54th American Society for Mass Spectrometry*, Seattle, Washington, May 28 - June 1, 2006.
 56. [E] PAN C, LANKFORD PK, ZHANG B, ODA Y, SAMATOVA NF, HARWOOD CS, PELLETIER DA, HETTICH RL, “Characterization of aromatic compound degradation pathways and their regulation in *Rhodospseudomonas palustris* using stable isotope labeling quantitative proteomics.” *Proceedings of the 54th American Society for Mass Spectrometry*, Seattle, Washington, May 28 - June 1, 2006.
 57. [CE] ZHANG Y, ABU-KHZAM FN, BALDWIN NE, CHESLE EJ, LANGSTON MA, SAMATOVA NF, “Genome-Scale Computational Approaches to Memory-Intensive Applications in Systems Biology.” *Proceedings of the ACM/IEEE Supercomputing 2005 Conference (SC 2005)*, pp. 12-12, 12-18 Nov 2005, Seattle, Washington, doi: 10.1109/SC.2005.29.
 58. [CE] LIN H, MA X, CHANDRAMOHAN P, GEIST A, SAMATOVA NF, “Efficient Data Access for Parallel BLAST.” *Proceedings of 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2005)*, pp. 72, 04-08 April 2005, doi: 10.1109/IPDPS.2005.190.
 59. [CE] YU G.X., PARK BH, CHANDRAMOHAN P, GEIST A, SAMATOVA NF, “In-silico prediction of surface residue clusters for enzyme-substrate specificity.” *Proceedings of the IEEE Computational Systems Bioinformatics Conference (CSB 2004)*, pp. 696-697, 16-19 Aug. 2004, doi: 10.1109/CSB.2004.1332548.
 60. [CE] YOGINATH S, SAMATOVA NF, BAUER D, KORA G, FANN G, GEIST A, “RScaLAPACK: High-performance parallel statistical computing with R and ScaLAPACK.” *Proceedings of the 18th International Conference on Parallel and Distributed Computing Systems (PDCS-2005)*, Sep 12-14, 2005, Las Vegas, Nevada.
 61. [M] ABU-KHZAM FN, BALDWIN NE, LANGSTON MA, SAMATOVA NF, “On the relative efficiency of maximal clique enumeration algorithms, with application to high-throughput computational biology.” *Proceedings of the International Conference on Research Trends in Science and Technology*, Beirut, Lebanon, March, 2005.
 62. [CE] SUTERS WH, ABU-KHZAM FN, ZHANG Y, SYMONS CT, SAMATOVA NF, AND LANGSTON MA, “A new approach and faster exact methods for the maximum common subgraph problem.” *Proceedings of the Computing and Combinatorics (COCOON)*, p. 717-727, 2005.
 63. [E] COX CD, MCCOLLUM JM, AUSTIN DW, DAR RD, ALLEN MS, SAMATOVA NF, SIMPSON ML, “Estimation of spectral properties and molecular biokinetic parameters from stochastic gene expression data.” *Proceedings of the Foundations of Systems Biology and Engineering (FOSBE)*, Santa Barbara, CA, August 2005.
 64. [M] THURMON BP, MCCOLLUM JM, PETERSON GD, COX CD, SAMATOVA NF, SAYLER GS, SIMPSON ML, “Accelerating exact stochastic simulation using reconfigurable computing.” *Proceedings of the Engineering of Reconfigurable Systems and Algorithms (ERSA)*, June 27-30, 2005, Monte Carlo Resort, Las Vegas, Nevada, USA.

65. [CE] PAN C, CHANDRAMOHAN P, PARK BH, MUNAVALLI RV, SAMATOVA NF, "Prediction of residue-residue contacts in domain interface by co-evolution." *Proceedings of the American Society for Microbiology 105th General Meeting*, Atlanta, GA, June 5, 2005.
66. [E] PAN C, HURST G, McDONALD H, PELLETIER D, SAMATOVA NF, HETTICH R, "Novel algorithms for extracting abundance ratios of stable isotope labeled peptide pairs from selected ion chromatograms in proteome samples." *Proceedings of the 53rd American Society for Mass Spectrometry*, San Antonio, Texas, June 2005.
67. [CE] PARK BH, OSTROUCHOV G, SAMATOVA NF, "Reservoir-based random sampling from data stream." *Proceedings of the Fourth SIAM International Conference on Data Mining*, Orlando, FL, April, 2004.
68. [E] OSTROUCHOV G, SAMATOVA NF, "Embedding methods and robust statistics for dimension reduction." *COMPSTAT 2004 Proceedings in Computational Statistics*, Physica-Verlag, A Springer Company, 2004, p359-370 (**Invited**).
69. [CE] PARK BH, MUNAVALLI R, GEIST A, SAMATOVA NF, "Analysis of interaction site predictions from separated data spaces." *Proceedings of the SIAM Bioinformatics Workshop (in conjunction with the Fourth SIAM International Conference on Data Mining)*, Orlando, FL, April, 2004.
70. [CE] YU GX, PARK BH, CHANDRAMOHAN P, GEIST A, SAMATOVA NF, "In-silico prediction of surface residue clusters for enzyme-substrate specificity." *Proceedings of the 2004 IEEE Computational Systems Bioinformatics (CSB2004)*, Stanford, CA, August 16-19, 2004, p. 696-697.
71. [M] PAN C, VERBERKMOES N, CHANDRAMOHAN P, SAMATOVA NF, HETTICH RL, "Integration of nanoscale HPLC-FTICRMS and HPLC-QIT for accurate mass measurements and high throughput MS/MS to achieve enhanced proteome characterization." *Proceeding of the 52nd American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics*, 2004.
72. [CL] SAMATOVA NF, YU GX, PARK BH, GEIST A, OSTROUCHOV G, "From genomics to functional proteomics: in silico approach." *A mini-symposium on Parallel Computational Biology (in conjunction with SIAM Conference on Parallel Processing for Scientific Computing*, February 25-27, 2004, San Francisco, CA (**Invited**).
73. [CE] YU, G.-X.; OSTROUCHOV, G.; GEIST, A.; SAMATOVA, N.F., "An SVM-based algorithm for identification of photosynthesis-specific genome features." *Proceedings of the IEEE Computational Systems Bioinformatics Conference (CSB 2003)*, pp. 235 - 243, 11-14 Aug. 2003, doi: 10.1109/CSB.2003.1227323, PMID: 16452798.
74. [CE] PARK, B.-H.; OSTROUCHOV, G.; YU, G.-X.; GEIST, A.; GORIN, A.; SAMATOVA, N.F., "Inference of protein-protein interactions by unlikely profile pair." *Proceedings of the Third IEEE International Conference on Data Mining (ICDM 2003)*, pp. 735-738, 19-22 Nov. 2003.
75. [CE] PARK, B.-H.; SAMATOVA, N.F., OSTROUCHOV, G.; GEIST, A., "Xmap: Fast dimension reduction algorithms for multivariate streamline data." *Proceedings of the 6th International Workshop on High Performance Data Mining: Pervasive and Data Stream Mining (in conjunction with Third International SIAM Conference on Data Mining)*, San Francisco, CA May 1-3, 2003.
76. [CE] ABU-KHZAM FN, SAMATOVA NF, OSTROUCHOV G, LANGSTON MA, GEIST GA, "Distributed dimension reduction algorithms for widely dispersed dataa." *Proceedings of the Fourteenth IASTED International Conference on Parallel and Distributed Computing and Systems (IASTED PDCS 2002)*, p. 167-174, 2002, ACTA Press.

77. [CL] SAMATOVA, N.F.; GEIST, A.; OSTROUCHOV, G.; MELECHKO, A.V., "Parallel out-of-core algorithm for genome-scale enumeration of metabolic systemic pathways." *Proceedings of the International Parallel and Distributed Processing Symposium (IPDPS 2002), the 1st Workshop on High Performance Computational Biology (HiCOMB2002)*, pp. 185-192, 15-19 April 2002, Fort Lauderdale, Florida,.
78. [E] POTOK, T.E.; ELMORE, M.T.; REED, J.W.; SAMATOVA, N.F., "An ontology-based HTML to XML conversion using intelligent agents." *Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS 2002)*, pp. 1220 - 1229, 7-10 Jan 2002.
79. [CL] SAMATOVA, N.F., POTOK, T.E.; LEUZE, M.R., "A vector perturbation approach to the generalized aircraft spare parts grouping." *Proceedings of the 10th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM 2002)*, pp. 1017-1026.
80. [CE] QU Y, OSTROUCHOV G, SAMATOVA NF, GEIST A, "Principal component analysis for dimension reduction in massive distributed data sets." *Proceedings of the Second SIAM International Conference on Data Mining*, p4-9, April 2002.
81. [E] POTOK, T.E.; IVEZIC N.E.; SAMATOVA, N.F., "Agent-based architecture for flexible lean cell design, analysis and evaluation." *Proceedings of the 4th Design of Information Infrastructure Systems for Manufacturing Conference (DIISM 2001)*, 2001, pp181-188, Melbourne, Victoria, Australia.
82. [CL] SAMATOVA NF, OSTROUCHOV G, GEIST A, MELECHKO AV, "RACHET: A new algorithm for mining multi-dimensional distributed datasets." *Proceedings of the SIAM Third Workshop on Mining Scientific Datasets*, Chicago, IL, April 2001.

Books and Book Chapters

83. [CL] SAMATOVA NF, BREIMYER P, KORA G, PAN P, YOGINATH S, “Parallel R for High Performance Analytics: Applications to Biology.” in *Scientific Data Management*, A. Shoshani and D. Rotem (editors), CRC Press/Taylor and Francis, 2008 (Coming soon). *New in 2008*
84. [CE] PARK BH, DAM P, PAN P, XU Y, GEIST A, HEFFELFINGER G, SAMATOVA NF, “In silico recognition of protein-protein interactions: theory and applications.” in *Advanced Data Mining Technologies in Bioinformatics*, Hui-Hwang Hsu (editor), Idea Group Inc., 2006.
85. [E] PULATOV AK, SAMATOVA NF, *Convex 3-polytopes*, Reference, Tashkent, Uzbekistan, 1992.

Patents

86. [E] POTOK TE, ELMORE MT, REED JW, TREADWELL JN, SAMATOVA NF, “System for gathering and summarizing internet information.” *US Patent Issued on July 4, 2006, Awarded on Jan 5, 2008* 07072883 Cl. 707-3. *New in 2008*

Invited Featured Articles

87. [CL] SAMATOVA NF, GORIN A, UBERBACHER E, KARPINETS T, PARK BH, PAN C, STRAATSMATP, CANNON W, RESAT H, LINS RD, OEHMEN C, “BioPilot: Data-driven computing for biological systems.” *SciDAC Review*, v. 5, p. 10-25, Fall 2007. *New in 2008*
88. [CL] SAMATOVA NF, SYMONS C, “Mentoring scientists of the future.” *SciDAC Review*, v. 5, p. 39-43, Spring 2007. *New in 2008*
89. [E] SHOSHANI A, CRICHLLOW T, ROSS R, SAMATOVA NF, “SDM Center: From data To discovery.” *SciDAC Review*, v. 2, p. 28-39, Fall 2006.

Invited Conference Proceedings

90. [CL] SAMATOVA, N.F., SCHMIDT, M., HENDRIX, W., BREIMYER, P., THOMAS, K., PARK, B.-H., “Coupling graph perturbation theory with scalable parallel algorithms for large-scale enumeration of maximal cliques in biological graphs.” *Journal of Physics: Conference Series – SciDAC 2008*, v. , July 13-17, 2008, Seattle, Washington, USA *New in 2009*
91. [CE] PARK BH, SAMATOVA NF, KARPINETS T, JALLOUK J, MOLONY S, HORTON S, ARCANGELI S, “Data-driven, data-intensive computing for modelling and analysis of biological networks: application to bioethanol production.” *Journal of Physics: Conference Series – SciDAC 2007*, v. 78, June 24-28, 2007, Boston, Massachusetts, USA *New in 2008*
92. [E] SHOSHANI, A., ALTINTAS, I., CHOUDHARY, A., CRITCHLOW, T., KAMATH, C., LUDSCHER, B., NIEPLOCHA, J., PARKER, S., ROSS, R., SAMATOVA, N.F., VOUK, M., “SDM Center Technologies for Accelerating Scientific Discoveries.” *Journal of Physics: Conference Series – SciDAC 2007*, v. 78, June 24-28, 2007, Boston, Massachusetts, USA *New in 2008*
93. [CL] SAMATOVA, N.F., BRANSTETTER, M., GANGULY, A.R., HETTICH, R., KHAN, S., KORA, G., LI, J., MA, X., PAN, C., SHOSHANI, A., S. YOGINATH, “High performance statistical computing with parallel R: Applications to biology and climate.” *Journal of Physics: Conference Series – SciDAC 2006*, v. 46, p. 505-509, 2006, Boston, Massachusetts, USA
94. [E] BETHEL W, ABRAM G, SHARF J, FRANK R, AHRENS J, SAMATOVA NF, MILLER M, “Interoperability of visualization software and data models is not an achievable goal.” In *Proceedings of the IEEE Visualization*, Seattle, Washington, October 19-24, 2003, p. 607-610.

Invited Workshop Research Roadmap Reports

95. [E] “Mathematics for Analysis of Petascale Data: Report on a Department of Energy Workshop,” Report from the Joint NSF/DOE Workshop, Editors: P. Kegelmeyer, R. Calderbank, R. Chritchlow, L. Jameson, C. Kamath, J. Meza, N.F. Samatova, A. Wilson, June 2008, <http://www.science.doe.gov/ascr/ProgramDocuments/PetascaleDataWorkshopReport.pdf>. *New in 2009*
96. [E] “Toward the Development of Predictive Theory and Modeling in Biology,” Report from the Joint ASCAC/BERAC Subcommittee on Modeling and Simulation for GTL, Editors: T. Zacharia, R. Stevens,..., N.F. Samatova, et. al., February 2008. *New in 2008*
97. [E] “Modeling and Simulation at the Exascale for Energy and the Environment,” DOE Report on the Advanced Scientific Computing Research Town Hall Meetings on Simulation and Modeling at the Exascale for Energy, Ecological Sustainability, and Global Security (E3), Editors: H. Simon, T. Zacharia, R. Stevens, January 2008. *New in 2008*
98. [M] “NSF Petascale Computing in the Biological Sciences,” Editors: A. Snavely, D.Bader, G. Jacobs, August 2006.
99. [E] “DHS Data Integration and Dissemination Workshop Report,” Editors: E. Stephen, N.F. Samatova, et. al, November 2005.
100. [E] “DHS Data Science Workshop Report,” Editors: T. Kolda, N. Samatova, et. al, September 2004.
101. [E] “DHS Advanced Scientific Computing Requirements Workshop Report,” Editors: S. Ashby,..., N.F. Samatova, et. al, October 2003.
102. [E] “DOE Office of Science Data-Management Workshops,” Editors: R. Mount,..., N.F. Samatova, et. al, March - May 2004.
103. [E] “NITRD High End Computing Revitalization Task Force (HECRTF): High End Computing for Full-Context Analysis and Visualization: When the Experiment is Done,” White paper by Ostrouchov G. and N.F. Samatova, June 2003.
104. [M] “DOE Visualization Frameworks Requirements Workshop,” Editors: R. Stevens, M. Papka,..., N.F. Samatova, et. al, June 2003.
105. [E] “DOE Report on Three Genomes to Life Workshops: Data Infrastructure, Modeling and Simulation, and Protein Structure Prediction,” Editors: T. Zacharia, R. Mann,..., N.F. Samatova, et. al, June 2003.
106. [L] “DOE OS Advanced Scientific Computing Research Accomplishments: Gleaning insight from scientific simulation data,” White paper by N.F. Samatova and Ostrouchov G., 2003.
107. [M] “DOE Report on the Mathematics Workshop for the Genomes to Life Program,” Editors: D. Brown, J. Guckenheimer, E. Ng,..., N.F. Samatova, et. al, March 2002.
108. [M] “DOE Report on the Computer Science Workshop for the Genomes to Life Program,” Editors: R. Bair, G. Johnson, J. Houghton, P. Karp, R. Stevens, B. Gropp,..., N.F. Samatova, et. al, March 2002.
109. [L] “DOE Report on the Computational Infrastructure Workshop for the Genomes to Life Program,” Editors: G. Heffelfinger, A. Geist,..., N.F. Samatova, et. al, January 2002.
110. [M] “DOE Report on the Computational Biology Workshop for the Genomes to Life Program,” Editors: M. Colvin, R. Mann, ..., Samatova NF, et. al., August, 2001

Technical Reports

111. [M] LIN H, MA X, LI J, SAMATOVA NF, TING Y, “Processor and Data Scheduling for Online Parallel Sequence Database Servers.” *August 2007: NCSU TR-2007-23.* *New in 2008*
112. [CE] LI J, MA X, YOGINATH S, KORA G, SAMATOVA NF, “Automatic, Transparent Runtime Parallelization of the R Scripting Language.” *January 2007: NCSU TR-2007-3.*
113. [CL] SAMATOVA NF, PARK BH, KRISHNAMURTHY R, MUNAVALLI R, SYMONS C, BUTTLER DJ, COTTOM T, CRITCHLOW TJ, SLEZAK T, “Information extraction from unstructured text for the Biodefense Knowledge Center.” *2005: United States. p. 18p., UCRL-CONF-213354.*
114. [E] SLEZAK T, CRITCHLOW T, HAZLETT S, SAMATOVA NF, CHANDRAMOHAN P, KRISHNAMURTHY R, “Design of the National Bioforensics library infrastructure.” *2004: United States. p. 22p., UCRL-TR-202217.*
115. [E] P. AGARWAL, J. AHRENS, D. BAUER, R. BLECK, P. CHANDRAMOHAN, T. DUNIGAN, G. FANN, M. GUEST, F. JAEGER, G. KORA, R. LATHAM, V. MEUNIER, L. OLIKER, R. ROSS, N. SAMATOVA, A. SHOSHANI, S. STUDHAM, B. SUMPTER, R. THAKUR, R. TOEDTE, J. S. VETTER, P. H. WORLEY, AND D. XU, “Evaluation of the SGI Altix 3700 at Oak Ridge National Laboratory.” *ORNL Technical Report, November, 2004.*
116. [E] BURRIS RD, CHOLIA S, DUNIGAN TH, FOWLER FM, GLEICHER MK, HOLMES HH, JOHNSTON NE, MEYER N, MILLION DL, OSTROUCHOV G, SAMATOVA NF, “Probe project status & accomplishments.” *2003, RNL/TM-2003/140.*

Tutorials

117. [CL] SAMATOVA, N.F., “Three-day Tutorial on Setting up a PC cluster for high performance computing.” Presented at the *Joint Institute for Computational Science (JICS) Workshop Series Summer, 1998, University of Tennessee.*
118. [CL] SAMATOVA, N.F., “Three-day Tutorial on Getting supercomputing power from a build-it-yourself PC cluster.” Presented at the *Joint Institute for Computational Science (JICS) Workshop Series Summer, 1999, University of Tennessee.*

Panels

1. “DOE OASCR Panel of Distinguished Members of the Computational Science Community to Identify Breakthroughs in Computational Science and Enabling Technologies across the Office of Sciences Scientific, Applied Mathematics, and Computer Science Research Programs,” March – June, 2008. *New in 2008*
2. “DOE Panel to Assess the Strategic Priorities of the Office of Advanced Scientific Computing Research Program,” January, 2008. *New in 2008*

Invited Research and Programmatic Presentations (Sole and Contributed)

1. “Coupling graph perturbation theory with scalable parallel algorithms for large-scale enumeration of maximal cliques in biological graphs,” Presented at the DOE SciDAC 2008 Conference, July 13-17, 2008, Seattle, WA. *New in 2009*
2. “Lessons learned from collaborations with GTL researchers,” Presented to the DOE ASCAC-BERAC Committee, October 4, 2007, San Francisco, CA. *New in 2008*

3. "Towards Uncovering Simplicity from Complexity in Biological Systems: Finding the Dots, Connecting the Dots, Understanding the Dots in Petascale Data," DOE Mathematics PI Meeting, May 22-24, 2007, Livermore, CA. *New in 2008*
4. "Scientific Data Management Center: Technologies, Applications, and Future Needs," DOE ASCR PI Meeting, Mar 31 – Apr 2, 2008, Denver, CO. *New in 2008*
5. "Scientific Data Management Center: Technologies and Applications," 2008 GTL-PI and ASCR-BER Bioenergy Research Center Workshop, Feb. 13, 2008, Rockville, MD. *New in 2008*
6. "The Shewanella Knowledgebase," Genomics:GTL Awardee Workshop VI and Metabolic Engineering Working Group, Interagency Conference on Metabolic Engineering, Bethesda, Maryland February 1013, 2008. *New in 2008*
7. "Comprehensive Integration of Regulatory Data in the Shewanella Knowledgebase," Genomics:GTL Awardee Workshop VI and Metabolic Engineering Working Group, Interagency Conference on Metabolic Engineering, Bethesda, Maryland February 1013, 2008. *New in 2008*
8. "The Center for Molecular and Cellular Systems: Biological Insights from Large Scale Protein-Protein Interaction Studies," Genomics:GTL Awardee Workshop VI and Metabolic Engineering Working Group, Interagency Conference on Metabolic Engineering, Bethesda, Maryland February 1013, 2008. *New in 2008*
9. "Computational Tools for Modeling of Biological Pathways and Networks," UT Health Science Center Meeting, March 6, 2008, Oak Ridge, TN. *New in 2008*
10. "Parallel, Scalable, Memory-Efficient Backtracking for Combinatorial Modeling of Large-Scale Biological Systems." The 22nd IEEE International Parallel and Distributed Processing Symposium (IPDPS 2008), the 7th IEEE International Workshop on High Performance Computational Biology (HiCOMB 2008), Miami, Florida USA, Apr 14-18, 2008. *New in 2008*
11. "Multi-stage Framework to Infer Protein Functional Modules from Mass Spectrometry Pull-Down Data with Assessment of Biological Relevance." IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2007), Nov 2-4, 2007. *New in 2008*
12. "The Buffered Work-Pool Approach for Search-Tree Based Optimization Algorithms." The 7th International Conference on Parallel Processing and Applied Mathematics (PPAM 2007), Gdansk, Poland, September 9-12, 2007. *New in 2008*
13. "Adaptive Request Scheduling for Parallel ScientificWeb Services." The 20th International Conference on Scientific and Statistical Database Management (SSDBM '08), Hong Kong, Jul 9-11, 2008. *New in 2008*
14. "Rapid and Robust Ranking of Text Documents in a Dynamically Changing Corpus." The 6th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA-08), Databases and Data Mining Track, Doha, Qatar, Mar 31 - Apr 4, 2008. *New in 2008*
15. "Scientific Data Management Center," Supercomputing Conference, November 10-16, 2007, Reno, NV. *New in 2008*
16. "BioPilot: Data-Intensive Computing for Complex Biological Systems," Supercomputing Conference, November 10-16, 2007, Reno, NV. *New in 2008*
17. "Uncovering Simplicity from Complexity: Discovering Fundamental Laws of Biology," UTK Mathematics Colloquium, September 20, 2007, Knoxville, TN. *New in 2008*
18. "Data Analytics at Petascale," UC-Davis Computer Science Seminar Series, March 13, 2007, Davis, CA.

19. "Data Analytics at Petascale: Accelerating Scientific Discovery," NCSU Computer Science Seminar Series, March 23, 2007, Raleigh, NC.
20. "Finding the Dots, Connecting the Dots, Understanding the Dots," WSU Computer Science Seminar Series, April 6, 2007, Pasco, WA.
21. "Common-target model for identification of protein interaction modules in pull-down experiments," Annual Meeting of the Institute for Biological Engineering, Mar 30-April 1, 2007, St. Louis, Missouri.
22. "Maximum common subgraph problem: faster solutions via vertex cover," ACS/IEEE International Conference on Computer Systems and Applications (AICCSA), May 13-16, 2007, Amman, Jordan (F. Abu-Khzam)
23. "Integrated knowledge resource for the Shewanella Federation," Joint Genomics:GTL Awardee Workshop V and Metabolic Engineering 2007 and USDA-DOE Plant Feedstock Genomics for Bioenergy Awardee Workshop, Feb 11-14, 2007, Bethesda, MD.
24. "Quantitative shotgun proteomics with ProRata: application to anaerobic aromatic degradation in *Rhodospseudomonas palustris*," Joint Genomics:GTL Awardee Workshop V and Metabolic Engineering 2007 and USDA-DOE Plant Feedstock Genomics for Bioenergy Awardee Workshop, Feb 11-14, 2007, Bethesda, MD.
25. "Scientific Data Management Center," DOE SciDAC Organization Worksho, Feb 5-6, 2007, Atlanta.
26. "Uncovering Mathematical Principles of Biology: Breaking the curses of dimensionality, intractability, & noise," presented to Dr. Michael Strayer, Acting Director of the DOE's Office of Advanced Scientific Computing (OASCR), March 6, 2006, ORNL.
27. "Comparative systems biology using scalable graph-theoretical approaches," MIT Computational Research Seminar, November 3, 2006 (host: Prof. Alan Edelman, MIT).
28. "Scalable graph-theoretical approaches to biological networks analysis," SIAM Conference on Parallel Processing for Scientific Computing, February 22-26, 2006, San Francisco, CA.
29. "Efficient data handling in comparative genome analysis applications," SIAM Conference on Parallel Processing for Scientific Computing, February 22-26, 2006, San Francisco, CA.
30. "High performance statistical computing with parallel R: applications to biology and climate modeling," The 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06), Nov 13-15, 2006, Washington, DC.
31. "Multi-criterion active learning in conditional Random Fields," DOE SciDAC Conference, June 25-29, 2006, Denver, CO.
32. "BioPilot: data-intensive computing for complex biological systems," UC Mercede Seminar Series, April 19, 2006, UC Mercede (host: Prof. Mike Colvin).
33. "Computational algorithms and software tools for quantitative shotgun proteomics," UC Mercede Seminar Series, April 19, 2006, UC Mercede (host: Prof. Mike Colvin).
34. "BioPilot: data-intensive computing for complex biological systems," Supercomputing Conference, November 11-17, 2006, Tampa, FL.
35. "Scientific Data Management Center," Supercomputing Conference, November 11-17, 2006, Tampa, FL.

36. "Comparative systems biology using scalable graph-theoretical approaches," Supercomputing Conference, November 11-17, 2006, Tampa, FL.
37. "Advanced technologies for identifying protein-protein interactions," Joint Genomics: GTL Contractor-Grantee Workshop IV and Metabolic Engineering Working Group Interagency Conference on Metabolic Engineering, Feb 12-15, 2006, Bethesda, MD.
38. "The Microbial Interactome Database: an online system for identifying interactions between proteins of microbial species," Joint Genomics: GTL Contractor-Grantee Workshop IV and Metabolic Engineering Working Group Interagency Conference on Metabolic Engineering, Feb 12-15, 2006, Bethesda, MD.
39. "Computational algorithms and software tools for quantitative shotgun proteomics," Joint Genomics: GTL Contractor-Grantee Workshop IV and Metabolic Engineering Working Group Interagency Conference on Metabolic Engineering, Feb 12-15, 2006, Bethesda, MD.
40. "Integrated knowledge resource for the Shewanella Federation," Joint Genomics: GTL Contractor-Grantee Workshop IV and Metabolic Engineering Working Group Interagency Conference on Metabolic Engineering, Feb 12-15, 2006, Bethesda, MD.
41. "Automatic parallelization for statistical computing with pR," DOE SciDAC SDM All Hands Meeting, Dec 11-13, 2006, Berkeley.
42. "Semantic annotation: how to make it painless, fast, and precise?," DHS Advanced Scientific Computing (ASC) All Hands Meeting, Oct 10-11, Livermore.
43. "Shewanella Federation knowledge integration resource," DOE Genomics:GTL Shewanella Federation Reverse Site Visit, September 18, Chantilly, VA.
44. "ProRata: a software package with improved point and interval estimation of protein abundance ratio for quantitative shotgun proteomics," American Society for Mass Spectrometry (ASMS) Conference, May 30 - April 2, 2006, Seattle, WA.
45. "Characterization of aromatic compound degradation pathways in *Rhodopseudomonas palustris* with stable isotope labeling quantitative shotgun proteomics and microarray," American Society for Mass Spectrometry (ASMS) Conference, May 30 - April 2, 2006, Seattle, WA.
46. "Genome-scale computational approaches to memory-intensive applications in systems biology," Supercomputing Conference, November 13-18, 2005 Seattle, WA.
47. "Systems biology framework for *R. palustris*," Environmental Microbes Meeting, September 15-16, 2005 Madison, WI (host: Prof. Julie Mitchell).
48. "Differential proteomics of *Rhodopseudomonas palustris* under its versatile metabolism states," International Conference on Microbial Genomes, September 15-16, 2005 Madison, WI.
49. "Novel algorithms for extracting abundance ratios of stable isotope labeled peptide pairs from selected ion chromatograms in proteome sample," American Society for Mass Spectrometry (ASMS) Conference, June 3 - 7, 2005, San Antonio, Texas.
50. "Statistical and graph theoretical approaches to semantic tagging of unstructured text for the Biodefence Knowledge Center," National Homeland Security R&D Conference, April 27-28, 2005, Boston.
51. "Statistical and graph theoretical approaches to semantic tagging of unstructured text for the Biodefence Knowledge Center," DHS TVTA Portfolio Offsite Conference, April 20-22, 2005, Las Vegas, NV.

52. "Advancing, integrating and deploying efficient statistical computing to high-throughput scientific applications," DOE SciDAC SDM All Hands Meeting, Oct 5-7, Raleigh, NC.
53. "Parallel R (pR) for high performance statistical computing," DOE SciDAC SDM All Hands Meeting, March 1-5, 2005, Salt Lake City, UT.
54. "Computer science and biology research at ORNL," Brown University Seminar, July 12-13, 2005 (hosts: Sam Fulkomer, Andy von Dam).
55. "Semantic tagging of unstructured text," DHS ASC PI Meeting, February 16, 2005, Livermore, CA.
56. "SciDAC II Plans," DOE SciDAC PI Meeting, Feb 17, 2005, Washington, DC (host: DOE OASCR Director, M. Strayer).
57. "Probabilistic weighting and semantic tagging of terms in free text," DHS TVTA Text Processing Group Meeting, June 20, 2005, Salt Lake City, Utah.
58. "Prediction of residue-residue contacts in domain interface by co-evolution," ASM 105th General Meeting, June 5, 2005, Atlanta, GA.
59. "Accelerating exact stochastic simulation using parallel supercomputing," Computational Methods in Systems Biology, April, 2005, Edinburgh, Scotland.
60. "A parallel implementation of Gillespie's exact stochastic simulation algorithm," International Symposium on Computational and Cellular Biology, March, 2005, Lenox, MA.
61. "RScaLAPACK: high-performance parallel statistical computing with R and ScaLAPACK," International Conference on Parallel and Distributed Computing Systems (PDCS-2005), September 12 - 14, 2005, Las Vegas, Nevada.
62. "The role of the SciDAC II in predictive DOE mission-driven systems biology," DOE SciDAC Conference, June 26-30, 2005, San Francisco, CA.
63. "Enabling microbial systems biology through advanced computing," Supercomputing Conference, November 13-18, Seattle, WA.
64. "Parallel R for high performance statistical computing," Supercomputing Conference, November 13-18, Seattle, WA.
65. "Computer science and biology research at ORNL," William & Marry University Seminar, May 15-16, 2005 (host: Prof. John von Rosendale).
66. "From biomolecular interactions to microbial transcriptional regulation of metabolism," DOE Sandia-ORNL Genomics:GTL Project Reverse Site Visit, May 5, 2005, Washington, DC.
67. "Integrating heterogeneous databases and tools for high throughput microbial analysis," GTL Contractor-Grantee Workshop, Feb 6-9, 2005, Washington, DC.
68. "Microbial Organism Encyclopedia," Sandia-ORNL Genomics:GTL Project All Hands Meeting, Apr 13, 2005, Seattle, WA.
69. "Protein-protein interaction prediction by examining co-occurrence of protein domains," Sandia-ORNL Genomics:GTL Project All Hands Meeting, Apr 13, 2005, Seattle, WA.
70. "From genomics to functional omics - in silico challenges and opportunities," NC State Genomic Sciences Seminar, March 23, 2005, Raleigh, NC (host: Prof. Xiasong Ma).
71. "Information extraction from unstructured text - concepts discovery, concepts mapping, relations inference," DHS ASC Program Review, Dec 9, 2005, Livermore, CA (host: DHS TVTA Program Manager, Sandy Landsburg).

72. "A new approach and faster exact methods for the maximum common subgraph problem," Computing and Combinatorics Conference (COCOON), August 16-19, 2005, Kunming, Yunnan.
73. "A parallel implementation of Gillespie's exact stochastic simulation algorithm," Third International Symposium on Computational Cell Biology, 2005, Lenox, MA.
74. "BioSpreadsheet: A biological model design, simulation, and analysis tool," Third International Symposium on Computational Cell Biology, 2005, Lenox, MA.
75. "Accelerating exact stochastic simulation using parallel supercomputing," International Conference on Computational Methods in Systems Biology (CMSB), 2005, Edinburgh, Scotland.
76. "Gene circuit inference by analysis of single cell reporter protein expression data," 6th Annual International Conference on Systems Biology, 2005, Boston, MA.
77. "Reservoir-based random sampling with replacement from a data stream," SIAM International Conference on Data Mining, April 22-24, 2004, Kissimmee, FL.
78. "Analysis of interaction site predictions from separated data spaces," SIAM International Conference on Data Mining, April 22-24, 2004, Kissimmee, FL.
79. "Surface patch ranking method identified cooperative substrate specificity residues in highly homologous enzymes," International Conference on Microbial Genomes, Sep 28-Oct 2, 2004 Durham, NC.
80. "In-silico prediction of surface residue clusters for enzyme-substrate specificity," IEEE Computational Systems Bioinformatics (CSB) Conference, Sep 28-Oct 2, 2004 Durham, NC.
81. "Surface patch ranking method identified cooperative substrate specificity residues in highly homologous enzymes," International Conference on Microbial Genomes, August 16-20, 2004, Stanford, CA.
82. "Data Intensive Analysis and Visualization," National Institute of Standards, May 6, 2004, Washington, DC.
83. "Embedding methods and robust statistics for dimension reduction," Symposium of the International Association for Statistical Computing (COMPSTAT-2004), August 23-27, 2004, Prague.
84. "Advancing data science to support DHS data intensive operational requirements," Supercomputing Conference, November 6-12, 2004, Pittsburgh, PA.
85. "Data intensive computing for complex biological systems," Supercomputing Conference, November 6-12, 2004, Pittsburgh, PA.
86. "Scientific Data Management Center - data mining and analysis," Supercomputing Conference, November 6-12, 2004, Pittsburgh, PA.
87. "SDM Center's Data Mining and Analysis," DOE SciDAC Conference, March 22-27, 2004, Charleston, SC.
88. "Systems level protein interactions annotation tools," Sandia-ORNL Genomics:GTL Project All Hands Meeting, Oct 20-21, 2004, San Diego, CA.
89. "An integrated computational environment for systems biology," Sandia-ORNL Genomics:GTL Project All Hands Meeting, Oct 20-21, 2004, San Diego, CA.
90. "Accelerating exact stochastic simulation using parallel supercomputing," DARPA PI Meeting, Oct 13, 2004, Vienna, VA.
91. "The Synechococcus Encyclopedia," Genomics:GTL Contractor-Grantee Workshop II, eb 29 - Mar 2, 2004, Washington, DC.

92. "Multi-resolution functional characterization of *Synechococcus* WH8102," Genomics:GTL Contractor-Grantee Workshop II, Feb 29 - Mar 2, 2004, Washington, DC.
93. "Parallel R: Intelligent task-level parallel execution in R," DOE SciDAC SDM All Hands Meeting, August 3-7, 2004, Berkeley, CA.
94. "PVTk: parallel netCDF reader and ROMIO geometry writer," DOE SciDAC SDM All Hands Meeting, August 3-7, 2004, Berkeley, CA.
95. "ASPECT Components," DOE SciDAC SDM All Hands Meeting, August 3-7, 2004, Berkeley, CA.
96. "Systems level protein interactions annotation tools," University of Georgia Systems Biology Seminar, Nov 21-22, 2004, Athens, GA (host: Prof. Ying Xu).
97. "From genomics to functional omics - in silico challenges and opportunities," SIAM minisymposium on parallel computational biology, February 26, 2004, San Francisco, CA.
98. "Statistical and graph theoretical approaches to semantic tagging of unstructured text for BKC," DHS Biodefence Knowledge Center (BKC) PI Meeting, February 22, 2004, Livermore, CA.
99. "Integration of HPLC-FTICR MS and HPLC-QIT MS2 to achieve enhanced proteome characterization," American Society for Mass Spectrometry (ASMS) Conference, May 23-27, 2004, Nashville, TN.
100. "An SVM-based algorithm for identification of photosynthesis-specific genomes features," IEEE Computational Systems Bioinformatics (CSB) Conference, August 11-14, 2003, Stanford, CA.
101. "Towards unraveling CO₂/O₂ specificity in microbial ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO)," Sandia-ORNL Genomics:GTL Project All Hands Meeting, Oct 21-22, 2003, Berkeley, CA.
102. "Identification of protein-protein interaction sites," Sandia-ORNL Genomics:GTL Project All Hands Meeting, Oct 21-22, 2003, Berkeley, CA.
103. "Bioinformatics tools for characterization of protein-protein interactions," Sandia-ORNL Genomics:GTL Project All Hands Meeting, March 26-27, 2003, Salt Lake, UT.
104. "PICUPP: Protein Interaction Classification by Unlikely Profile Pair," Sandia-ORNL Genomics:GTL Project All Hands Meeting, March 26-27, 2003, Salt Lake, UT.
105. "Xmap: fast dimension reduction algorithms for multivariate streamline data," SIAM Conference on Data Mining, May 1-3, 2003, San Francisco, CA.
106. "Distributed data mining," Spring Research Conference on Statistics, Analysis and Visualization of Massive Simulation Data Sets, June 4-6, 2003, Dayton, Ohio.
107. "Can dimension reduction be fast and robust? FastMap and the Convex Hull of multivariate data," UTK Statistics Seminar, December 6, 2003, UTK.
108. "Interoperability of visualization software and data models is Not an achievable goal," IEEE Visualization, October 19-24, 2003, Seattle, Washington.
109. "Gleaning insight from scientific simulation data," DOE SciDAC Program Review, March 12, 2003, Washington DC.
110. "Advanced data science for homeland security," Supercomputing Conference, November 15-21, 2003, Phoenix.
111. "ASPECT: Adaptable Simulation Product Exploration and Control Tool," Supercomputing Conference, November 15-21, 2003, Phoenix.

112. "Sandia-ORNL Genomes to Life Project," Supercomputing Conference, November 15-21, 2003, Phoenix.
113. "Analysis of protein complexes from a fundamental understanding of protein binding domains and protein-protein interactions in *Synechococcus* WH8102," DOE Genomes to Life Contractor - Grantee Workshop I, Feb 9-12, 2003, Arlington, Virginia.
114. "Carbon sequestration in *Synechococcus* sp.: from molecular machines to hierarchical modeling," DOE Genomes to Life Contractor - Grantee Workshop I, Feb 9-12, 2003, Arlington, Virginia.
115. "Data exploration environment," DOE Scientific Data Management Workshop, Aug 28, 2003, Argonne, IL (host: DOE Program Manager, John von Rosendale, Arie Shoshani, LBL).
116. "Adaptable Simulation Product Exploration and Control Tool (ASPECT)," SciDAC Terascale Supernova Initiative All Hands Meeting, February, 2002 (host: Dr. Tony Mezzacappa).
117. "RACHET: petascale distributed data analysis suite," SPEEDUP Workshop on Distributed Supercomputing Data Intensive Computing, March 4-6, 2002, Leukerbad, Valais, Switzerland.
118. "CADS challenges in systems biology: a computer science perspective," NCSA Alliance All-Hands Meeting, May 8-10, 2002, Urbana.
119. "Parallel out-of-core algorithm for genome-scale enumeration of metabolic systemic pathways," First International Workshop on High Performance Computational Biology (HiCOMB), International Parallel and Distributed Processing Symposium (IPDPS-02), April 15, 2002, Fort Lauderdale, Florida.
120. "Advanced algorithms for computational biology," DOE Science and Technology Review, February 11, 2002, Oak Ridge, TN.
121. "Scientific Data Management Center," DOE Program Review, August 16, 2002, ORNL (presentation to DOE SciDAC program director, Dr. Alan Laub).
122. "Advanced algorithms for computational biology," DOE Program Review, August 16, 2002, ORNL (presentation to DOE OS program manager, Dr. Walter Stevens).
123. "ASPECT: Adaptable Simulation Product Exploration and Control Toolkit," SciDAC SDM All Hands Meeting, September 11-13, 2002, San Diego.
124. "Data mining and access pattern discovery," DOE SciDAC Conference, June 26-27, 2002, Argonne.
125. "High resolution tools and algorithms for genome scale analysis," Supercomputing Conference, Nov 16-21, 2002, Baltimore, MD.
126. "ASPECT - data stream monitoring tool," Supercomputing Conference, Nov 16-21, 2002, Baltimore, MD.
127. "Research in algorithms for scalable analysis of distributed and streaming data," Supercomputing Conference, Nov 16-21, 2002, Baltimore, MD.
128. "Medard W. Welch Award Lecture: Intertwined charge density wave and defect-ordering phase transitions in a 2-D system," IUUSTA 15th International Vacuum Congress (IVC-15), AVS 48th International Symposium (AVS-48), 11th International Conference on Solid Surfaces (ICSS-11), San Francisco, October 2002(invited talk by Prof. W. Plummer, UTK).
129. "Combining distributed local principal component analyses into a global analysis," C. Warren Neel Conference on Statistical Data Mining and Knowledge Discovery, June 22-25, 2002, Knoxville, TN.

130. "Annotation of DNA regulatory regions through the utilization of underlying recognition principles," C. Warren Neel Conference on Statistical Data Mining and Knowledge Discovery, June 22-25, 2002, Knoxville, TN.
131. "Multivariate analysis of massive distributed data sets," Spring Research Conference on Statistics in Industry and Technology, May 20-22, 2002, Ann Arbor, Michigan.
132. "Evolutionary analysis of enzymes," International Meeting on Microbial Genomes, September 8-12, 2002, Lake Arrowhead, CA.
133. "Resource and Location Aware Mining: Principal component analysis for dimension reduction in massive distributed data sets," 5th International Workshop on High Performance Data Mining (HPDM), April 13, 2002, Washington, DC.
134. "RACHET: a new algorithm for mining multi-dimensional distributed dataset," SIAM Third Workshop on Mining Scientific Datasets, Chicago, IL, April 5-7, 2001.
135. "Characterization of the solution space from metabolic flux balance analysis," UCSD Genetic Circuits Group Seminar, July 12, 2001 (host: Prof. Bernhard Palsson, UCSD).
136. "Multi-agent based high-dimensional cluster analysis," DOE Program Review, August 9, 2001 (presentation to DOE OS program manager, Dr. Steve Eckstrand).
137. "Multi-agent based high-dimensional cluster analysis," Supercomputing Conference, July 10, 2001, Gatlinburg, TN.
138. "Cracking computational complexity for genome-scale modeling of biochemical pathways," SciDAC SDM Kick-off Meeting, November 10-16, 2001, Denver, CO.
139. "RACHET: petascale distributed data analysis suit," SciDAC SDM Kick-off Meeting, November 10-16, 2001, Denver, CO.
140. "Predictive models of biochemical pathways and microbial behavior," DOE Genomes to Life (GTL) Workshop, August 7, 2001 (invited talk by Prof. B. Palsson, UCSD).

Organizing Committees

1. CO-CHAIR, *DOE/NSF Mathematics for Analysis of Petascale Data*, June 3–5, 2008, Rockville, Maryland *New in 2008*
2. CO-CHAIR, *DOE Genomics:GTL Program Systems Biology Network/Knowledgebase Workshop*, May 28 - May 31, 2008, Bethesda North, Washington D.C. *New in 2008*
3. CO-CHAIR, *DOE/OASCR Mathematics for Peta-scale Data Analysis & Computation*, June 2007 *New in 2008*
4. TECHNICAL SESSION CHAIR, *DOE Workshop on Modeling and Simulation at the Exascale for Energy and the Environment*, Apr 17 - 18, 2007, San Francisco, CA. *New in 2007*
5. TECHNICAL SESSION CHAIR, *DOE Workshop on Modeling and Simulation at the Exascale for Energy and the Environment*, May 17 - 18, 2007, Oak Ridge, TN. *New in 2007*
6. TECHNICAL SESSION CHAIR, *DOE Workshop on Modeling and Simulation at the Exascale for Energy and the Environment*, May 31 - June 1, 2007, Argonne, IL. *New in 2007*
7. COMMITTEE MEMBER, *DOE Data Sharing & Integration Summit*, April 20, 2006, Berkeley.
8. COMMITTEE MEMBER, *DOE SciDAC Conference*, June 26-30, 2005, San Francisco, CA.
9. CHAIR, *DOE SciDAC Biology Summit*, August 8-9, 2005, Reston, VA.

10. CO-CHAIR, *DHS Threat Awareness Customer Data Integration & Dissemination*, November 1-2, 2005, Alexandria VA.
11. CO-CHAIR, *DHS Data Sciences for Homeland Security Information Management and Knowledge Discovery*, Sep. 22-23, 2004, Alexandria, VA.
12. COMMITTEE MEMBER, *Scientific Data Management*, April 20-22, Menlo Park, CA.
13. COMMITTEE MEMBER, *DOE Computational Infrastructure for the Genomes to Life Program*, July 22, 2002, Gaithersbrug, MD.

Invited Workshop Participant

1. PARTICIPANT, *Extremely Large Scale Databases (XLDB) Workshop*, Oct. 25, 2007, Stanford, CA. *New in 2008*
2. PARTICIPANT, *2008 GTL-PI and ASCR-BER Bioenergy Research Center Workshop*, Feb. 13, 2008, Rockville, MD. *New in 2008*
3. PARTICIPANT, *DOE OASCR PI Meeting*, Mar 31 – Apr 2, 2008, Denver CO. *New in 2008*
4. PARTICIPANT, *NSF Petascale Computing in the Biological Sciences*, Aug. 29-30, 2006.
5. SESSION CHAIR, *Workshop on Enabling Petascale Science and Engineering Applications*, Dec 9, 2005, GA Tech, Atlanta.
6. SESSION CHAIR, *DOE SciDAC II Planning Workshop on Accelerating Scientific Discovery in Experimental Science via Advanced Computing*, Aug 2-3, 2005, Argonne, IL.
7. SESSION CO-CHAIR, *DOE BER/ASCR GTL Deep Dive Technology*, June 14-16, 2004, Alexandria, VA.
8. SESSION CO-CHAIR, *DOE Scientific Data Management*, March 16-18, 2004, Menlo Park, CA.
9. PARTICIPANT, *DOE Scientific Data Management*, March 16-18, 2004, Menlo Park, CA.
10. PARTICIPANT, *DOE Scientific Data Management*, May 24-26, 2004, Chicago, IL.
11. SESSION CO-CHAIR, *DHS Advanced Scientific Computing Requirements*, Oct 8-9, 2003, Arlington, VA.
12. PARTICIPANT, *DOE/OBER GTL and Beyond Data Standards*, Sep 10-11, 2003, Berkeley, CA.
13. PARTICIPANT, *NITRD High End Computing Revitalization Task Force (HECRTF)*, June 15-18, 2003, Washington, DC.
14. PARTICIPANT, *DOE Visualization Frameworks Requirements*, June 2-3, 2003, Bethesda, MD.
15. PARTICIPANT, *DOE Neutron Science Software Initiative (NeSSI)*, Oct 13-15, 2003, Oak Ridge, TN.
16. PARTICIPANT, *DOE Genomes To Life Data Infrastructure*, July 22, 2003, Gaithersburg, MD.
17. PARTICIPANT, *DOE Mathematics for the Genomes-to-Life Program*, March 18-19, 2002, Gaithersburg, MD.
18. PARTICIPANT, *DOE Computer Science for GTL*, March 6-7, 2002, Gaithersburg, MD.
19. PARTICIPANT, *DOE First Genomes to Life (GTL) Computational Biology Workshop*, August 7-8, 2001.

Conference Program Committees

1. PC MEMBER, *The 2007 IEEE International Symposium on Bioinformatics and Life Science Computing (BLSC)*, in conjunction with *The IEEE 21st International Conference on Advanced Information Networking and Applications (AINA-07)*, ay 21-23, 2007, Niagara Fall, Ontario, Canada.
2. PC MEMBER, *2nd IEEE Workshop on High Performance Computing in Medicine and Biology (HiPCoMB-2006)*, in conjunction with the *12th International Conference on Parallel and Distributed Systems (ICPADS 2005)*, April 18-20, Vienna, Austria.
3. PC MEMBER, *1st IEEE Workshop on High Performance Computing in Medicine and Biology (HiPCoMB-2005)*, in conjunction with the *11th International Conference on Parallel and Distributed Systems (ICPADS 2005)*, July 20-22, 2005, Fukuoka, Japan.
4. PC MEMBER, *16th International Conference on Scientific and Statistical Database Management (SSDBM)*, June 21-23, 2004, Santorini Island, Greece.
5. PC MEMBER, *Third SIAM International Conference on Data Mining, Workshop on Data Mining for Counter Terrorism and Security*, May 3, 2003, San Francisco, CA.
6. PC MEMBER, *IEEE Bioinformatics Conference*, August 11-14, 2003, Stanford, CA.
7. TECHNICAL PROGRAM CO-CHAIR, *C. Warren Neel Conference on Statistical Data Mining*, June 22-25, 2002, Knoxville, TN.
8. PC MEMBER, *16th International Conference on Scientific and Statistical Database Management (SSDBM)*, April 23-27, 2002, Baltimore, MD.
9. PC MEMBER, *CLUSTER: IEEE International Conference on Cluster Computing*, September 23-26, 2002, Chicago, IL.

B. Research Project Record

Pending Grants

1. **Title:** *Ultrascale Computational Modeling of Phenotype-Specific Metabolic Processes in Microbial Communities* *New in 2009*
Agency/Program: DOE
PI: Nagiza F. Samatova
co-PIs: Jill Banfield (UC Berkeley), Robert Hettich (ORNL), Pan Chongle (ORNL)
Period: 10/2008 - 9/2011
Amount: \$1,500,000
In-coming to NCSU Amount: \$450,000
Status: Under review at DOE, submitted on July 28th, 2008.

Awarded Grants

1. **Title:** *Scientific Data Management Technologies to Accelerate Fusion Scientific Discovery: A Strategic Application Partnership (SAP) Proposal* *New in 2008*
Agency/Program: DOE
PIs: Arie Shoshani (LBL), Nagiza F. Samatova, Scott Klasky
Period: 07/2008 - 03/2011
Amount: \$600,000
In-coming to NCSU Amount: \$150,000
Status: **Approved** for funding.
Comment: NCSU funding will be delivered as a subcontract from ORNL.

Current Sponsored Grants

1. **Title:** *Data Intensive Computing for Complex Biological Systems* *New in 2008*
Agency/Program: DOE
PIs: Nagiza F. Samatova and Tjerk Straatsma (PNNL)
Period: 10/2007 - 9/2009
Amount: \$1,500,000
In-coming to NCSU Amount: \$400,000
2. **Title:** *Workshop on Mathematics for Petascale Data* *New in 2008*
Agency/Program: NSF
PI: Nagiza F. Samatova
Period: 5/2008 - 9/2008
Amount: \$20,000 (NCSU)
3. **Title:** *Data Intensive Computing for Complex Biological Systems* *New in 2008*
Agency/Program: DOE
PIs: Nagiza F. Samatova and Tjerk Straatsma (PNNL)
Period: 9/2004 - 9/2007
Amount: \$2,700,000
In-Coming to NCSU Amount: \$200,000
Comment: The NCSU funding was recorded in RADAR under the ORNL-NCSU contracts 'High-Performance Data Analytics with Demonstrations to DOE-Mission Applications'

4. **Title:** *DOE SciDAC Scientific Data Management Center* *New in 2008*
Agency/Program: DOE
ORNL PI: Nagiza F. Samatova
Period: 7/2006 - 6/2011
Amount: \$2,700,000
In-coming to NCSU Amount: \$1,500,000
Comments: \$520,000 out of \$1,500,000 NCSU funding has been recorded in RADAR under the ORNL-NCSU contract 'High-Performance Data Analytics with Demonstrations to DOE-Mission Applications' (Current). Only ORNL funding is reflected. The overall SciDAC SDM Center (PI: Arie Shoshani, LBL) budget is not reflected here.
5. **Title:** *Integrated Knowledge Resource for the Shewanella Federation* *New in 2008*
Agency/Program: DOE
PIs: Edward Uberbacher (ORNL) and Nagiza Samatova
Period: 10/2005 - 9/2008
Amount: \$3,300,000
In-coming to NCSU Amount: \$228,000
Comment: The NCSU funding is recorded under the ORNL-NCSU contract 'Joint Faculty Agreement' (Current).
6. **Title:** *Integrated Genome-Based Studies of Shewanella Ecophysiology* *New in 2008*
Agency/Program: DOE
PIs: Edward Uberbacher (ORNL), Nagiza Samatova, etc.
Period: 10/2006 - 9/2009
Amount: \$1,100,000
In-coming to NCSU Amount: \$80,000
Comment: The NCSU funding is recorded under the ORNL-NCSU contract 'Joint Faculty Agreement' (Current). The overall Shewanella Federation (PI: Jim Fredrickson, PNNL) budget is not reflected here. Only the portion of the funding for which the responsibilities of the PIs are counted are reflected.
7. **Title:** *Modeling Cellular Mechanisms for Efficient Bioethanol Production through Petascale Comparative Analysis of Biological Networks* *New in 2008*
Agency/Program: UT-Battelle ORNL/LDRD
PIs: Andrey Gorin (ORNL) and Nagiza Samatova
Period: 10/2007 - 9/2008
Amount: \$350,000
In-coming to NCSU Amount: \$50,000
Comment: The NCSU funding is recorded under the ORNL-NCSU contract 'Joint Faculty Agreement' (Current).

Past Sponsored Grants

1. **Title:** *Modeling Cellular Mechanisms for Efficient Bioethanol Production through Petascale Comparative Analysis of Biological Networks*
Agency/Program: UT-Battelle ORNL/LDRD
PIs: Nagiza Samatova
Period: 10/2006 - 9/2007
Amount: \$400,000 (ORNL)

2. **Title:** *DOE SciDAC Scientific Data Management Center*
Agency/Program: DOE
ORNL PI: Nagiza F. Samatova
Period: 7/2001 - 6/2006
Amount: \$2,700,000 (ORNL)
Comments: Only ORNL funding is reflected. The overall SciDAC SDM Center (PI: Arie Shoshani, LBL) budget is not reflected here.
3. **Title:** *Biodefense Knowledge Center (BKC)*
Agency/Program: DHS/TVTA
PIs: Nagiza Samatova
Period: 04/2006 - 03/2007
Amount: \$650,000
4. **Title:** *Natural Language Processing for Biological Text Mining*
Agency/Program: DHS/ASC
PIs: Nagiza Samatova
Period: 10/2003 - 09/2006
Amount: \$800,000
5. **Title:** *Large Scale Exploration of Protein Models for System Biology Applications*
Agency/Program: UT-Battelle ORNL/LDRD
PIs: Andrey Gorin and Nagiza Samatova
Period: 10/2005 - 9/2007
Amount: \$385,000 (ORNL)
6. **Title:** *Systems biology framework for postgenomic microbiology*
Agency/Program: UT-Battelle ORNL/LDRD
PIs: Edward Uberbacher, Andrey Gorin, and Nagiza Samatova
Period: 10/2004 - 9/2006
Amount: \$520,000 (ORNL)
7. **Title:** *National Bio-forensics Information Encyclopedia: Heterogeneous Data Integration*
Agency/Program: DHS/ASC
PIs: Nagiza Samatova
Period: 10/2002 - 09/2003
Amount: \$250,000
8. **Title:** *Carbon Sequestration of Synechococcus: From Molecular Machines to Hierarchical Modeling*
Agency/Program: DOE/OASCR
PIs: Grant Heffelfinger (SNL), Al Geist (ORNL)
co-PIs: Nagiza Samatova, Andrey Gorin and others)
Period: 10/2002 - 09/2005
Amount: \$9,000,000
9. **Title:** *SKALE: Scalable Clustering and Feature Extraction from Massive High-Dimensional Distributed Scientific Data (under the Probe project)*
Agency/Program: DOE
PIs: Nagiza Samatova
Period: 01/2000 - 12/2003
Amount: \$900,000

10. **Title:** *High-throughput Biological Data Analysis and Modeling Tools for Genomes To Life Facilities*
Agency/Program: UT-Battelle ORNL/LDRD
PIs: Nagiza Samatova
Period: 10/2002 - 09/2004
Amount: \$500,000 (ORNL)
11. **Title:** *Biologically Driven Controlled Synthesis and Directed Assembly of Nanophase Inorganic Materials*
Agency/Program: UT-Battelle ORNL/LDRD
PIs: Michael Simpson (ORNL) and Nagiza Samatova
Period: 10/2003 - 09/2005
Amount: \$350,000 (ORNL)
12. **Title:** *Scalable Tools for Petascale Distributed Data Analysis*
Agency/Program: UT-Battelle ORNL/LDRD
PIs: George Ostrouchov (ORNL) and Nagiza Samatova
Period: 10/2001 - 09/2003
Amount: \$630,000 (ORNL)
13. **Title:** *PDQ: A PACI Petascale Data Quest*
Agency/Program: NSF
PIs: Nagiza Samatova
Period: 06/2002-05/2005
Amount: \$150,000
Comments: Only ORNL funding is reflected. The overall project (PI: Ian Foster, ANL) budget is not reflected here.
14. **Title:** *Scientific Workspaces of the Future: Grid-based Visualization and Collaboration Services*
Agency/Program: NSF
PIs: Nagiza Samatova
Period: 06/2002-05/2004
Amount: \$100,000
Comments: Only ORNL funding is reflected. The overall project (PI: Rick Stevens, ANL) budget is not reflected here.
15. **Title:** *Program Development for Computational Biology*
Agency/Program: UT-Battelle ORNL/LDRD
PIs: Nagiza Samatova
Period: 10/2001 - 09/2006
Amount: \$60,000 (ORNL)

C. Participation in Centers, Consortia and Other Organized Scholarly Efforts

JOINT FACULTY APPOINTEE, Oak Ridge National Laboratory (ORNL), 2007-present

ASSOCIATE FACULTY MEMBER, Center for High Performance Simulation (CHiPS), 2007-present

AFFILIATED FACULTY, UT-ORNL Graduate School of Genome Science and Technology, 2002-present

RESEARCH AFFILIATE, UT-ORNL Joint Institute for Computational Sciences (JICS), 1997-present

AREA LEAD, DOE Scientific Data Management Center, 2001-present

AREA CO-LEAD, DOE Shewanella Federation, 2004-present

DIRECTOR, DOE BioPilot Center for Data Intensive Computing in Biology, 2004-present

AREA LEAD, DOE Sandia-ORNL Genomics:GTL Center, 2002-2004

D. Cross-Disciplinary Activities

CO-ADVISOR, PhD student, Chongle Pan, with Bob Hettich (ORNL, Analytics Chemistry), 2004-2006

CO-ADVISOR, PhD student, Heshan Lin, with Xiasong Ma (NCSU, Systems), 2004-present

CO-ADVISOR, PhD student, Jiangtin Li, with Xiasong Ma (NCSU, Systems), 2005-present

CO-ADVISOR, PhD student, Andrea Rocha, with Peter Stroot (USF, Biology), 2006-present

IV. Extension and Engagement with Constituencies Outside the University

A. Media Coverage

THE NEWS AND OBSERVER

1. 'Six who live or work in OR named YWCA Tribute to Women finalists,' Oak Ridger, August 2, 2007 *New in 2008*
2. 'Outstanding mentors and teachers honored at ORNL,' ORNL News Release, Feb. 14, 2008 *New in 2008*
3. 'U.S. Patents Awarded to Inventors in Tennessee,' Targeted News Service, Jan 5, 2008 *New in 2008*
4. 'Rising Stars,' Science 314, 1665e (2006)
5. 'National Competitive Investment Act, Senator Mr. Alexander, The Library of Congress, Congressional Record, Senate, Page S11588, Dec. 8, 2006
6. 'New Opportunities for Data and Information: Finding the Dots, Connecting the Dots, Understanding the Dots,' Presentation by Dr. Raymond L. Orbach, Director, Office of Science, at The Expanding Universe of Digital Data Collections Symposium, 2006 AAAS Annual Meeting, St. Louis, MO
7. Nation's Top Science Prize for High School Students Goes to Mathematics and Bioinformatics Research,' 2006-07 Siemens Competition in Math, Science & Technology News Release, Dec 6, 2006
8. 'ProRata TOOLbox,' Journal of Proteome Research, Vol. 5, No. 11, 2006
9. 'From Data to Discovery, SciDAC Review, Fall 2006
10. 'Gu semifinalist in Intel Science Talent Search,' Oak Ridger, January 24, 2007
11. 'Oak Ridge: Science making history,' Oak Ridger, January 23, 2007
12. 'ORHS team wins \$100K,' Oak Ridger, December 5, 2006
13. 'Oak Ridge's brightest on top,' Knoxville News Sentinel, December 5, 2006
14. '3 OR students: National science champs,' Knoxville News Sentinel, December 4, 2006
15. 'ORHS team wins regional Siemens contest,' Oak Ridger, November 9, 2006
16. Oak Ridge High students Siemens Competition finalists,' Oak Ridger, November 1, 2006
17. Local students take home \$100,000 prize in national science competition,' Wate 6, The News Radio Station, December 4, 2006
18. 'Massachusetts & Tennessee Teens Take Regional Title in Nation's Top High School Science Competition,' Siemens News Release, November 4, 2006
19. 'ORNL-Mentored Oak Ridge High School team wins Siemens Competition,' Oak Ridge, Dec.6, 2006
20. 'High School Bioenergy Champions' ORNL Review, 2007
21. 'Tennessee students and researcher win \$900,000 US for biofuel mathematical research (R&D),' Bioenergy and Biofuels, Dec 11, 2006

22. 'Oak Ridge's brightest on top,' Innovation Valley - Nanotech, December 5, 2006
23. 'Siemens Science & Technology Competition National Winners,' Cogito, December 5, 2006
24. 'Oak Ridge High School Team wins \$100K,' Rarity Community, December 5, 2006
25. 'Scientists' 'genetic noise' research published in Nature,' Oak Ridger, February 7, 2006
26. 'Gene Network Shaping of Inherent Noise Spectra,' Center for Nanophase Materials Center News, 2006
27. 'ORNL supercomputer helps trio excel, Computer Science and Mathematics PR, ORNL, Dec 5, 2006
28. 'Last-minute idea collects dividends,' Advanced Physics Forums, Nov 7, 2005
29. 'Nation's Top Science Prize for High School Students Goes to Mathematics and Bioinformatics Research,' Yahoo News, Dec 4, 2006
30. 'Nation's Top Science Prize for High School Students Goes to Mathematics and Bioinformatics Research,' Digital50, Dec 4, 2006
31. 'UT and ORNL scientists' genetic research published in journal,' The Oak Ridge Observer, Feb 9, 2006
32. 'Oak Ridge High School students to participate in national science competition,' Oak Ridge, TN, Dec. 6, 2005
33. 'ORNL leading effort to help harness power of Shewanella,' Oak Ridge, TN, Nov 11, 2005
34. 'ORHS team wins \$30,000 at Siemens Westinghouse competition,' Oak Ridger, December 6, 2005
35. 'ORHS team wins regional science competition, headed to nationals,' Oak Ridger, November 23, 2005
36. 'ORNL leading research efforts involving Shewanella microbe,' Oak Ridger, November 18, 2005
37. '3 from ORHS win regional science event,' Computer Science and Mathematics PR, ORNL, Nov 21, 2005
38. 'Three from ORHS Win Regional Science Event,' Tennessee Valley Corridor, November 11, 2005
39. 'Project leads students to high-profile academic competition,' Oak Ridge Observer, Issue 48, Vol 1, 2005
40. 'New Biological Discoveries through Data Intensive Computing,' Linux Electrons, Nov 13, 2004
41. 'New Biological Discoveries through Data Intensive Computing,' Supercomputing Online, Nov 3, 2004
42. 'Two DOE Labs Buck the Biocluster Trend to Test-Drive a Few Architecture Alternatives,' GenomeWeb, Volume 8, Number 47, December 6, 2004
43. 'HPC Beyond Clusters,' John S. McNeil, Genome Technology, March, 2004
44. 'Data intensive computing leads to biological discoveries,' Oak Ridge, TN, Dec 10, 2004
45. 'Computational Technology,' CCN Magazine, Dec 14, 2004
46. 'ORNL staffers honored for support of tech transfer initiatives,' Oak Ridger, May 29, 2003

47. 'Bandwidth Challenge Teams Push Networking Performance Envelope at SC2003 Conference - Sustained 23 Gigabits per Second Sets New Record,' Supercomputing 2003, Application Foundation Award
48. 'Bandwidth Challenge Teams Push Networking Performance Envelope at SC2003,' Supercomputing Online, Nov 25, 2003
49. 'Bandwidth Challenge Teams Push Networking Performance Envelope at SC2003,' Daily News and Information for the Global Grid Community, December 8, 2003: Vol 35, No 49
50. 'Developing Computer Tools for Scientists,' ORNL Review, Vol. 35, No. 1, 2002
51. 'Developing computer tools for scientists,' DOE OS Research News, 2002
52. 'Retaining and Retrieving Data More Effectively,' ORNL Review, Vol. 35, No. 1, 2002
53. 'Retaining and Retrieving Data More Effectively,' DOE OS Research News, 2002

B. Public Service

- OAK RIDGE HIGH SCHOOL MATHEMATICS THESIS PROGRAM

During 2005-2007 I have mentored more than half a dozen senior high school students in their Mathematics Thesis Program. The students participation in this program resulted in a number of national level students' awards (see 'High School Math and Sciences Theses Directed' section)

New in 2008

- DOE RESEARCH ALLIANCE IN MATH AND SCIENCES (RAMS) PROGRAM

Since 2003, each summer I mentor two-to-three internship students from Black Colleges and Universities and Minority Educational Institutions.

New in 2008

- SUMMER STUDENTS TRAINING PROGRAMS Since 2005, each summer I teach various 1-4 week workshops for senior high school and undergraduate students:

Computational Systems Biology, 2007

Enrollment: 10 undergraduate from UC Mercede and 4 high school students from ORHS

New in 2008

Object-Oriented Programming, 2006

Statistical Theory: Basics, 2005-2006

Graph Theory and Algorithms, 2005-2006

Artificial Intelligence Methods, 2005

V. Technological and Managerial Innovation

A. Technology Transfer

1. VIPAR: VIRTUAL INFORMATION PROCESS RESEARCH AGENT, System for gathering and summarizing internet information.
Impact: US Patent, Awarded on Jan 5, 2008 07072883 Cl. 707-3. *New in 2008*
2. SHEWANELLA KNOWLEDGEBASE, Integrated Knowledgebase for DOE Shewanella Federation.
Impact: A continuously growing knowledgebase used by a large community of users. *New in 2008*
3. PRODATA: QUANTITATIVE SHOTGUN PROTEOMICS SOFTWARE, Open source software distributed from: <http://www.MSProData.org>
Impact: More than 1000s downloads since June 2006 till present; featured in the *Journal of Proteome Research*, Vol. 5, No. 11, 2006.
4. PR: PARALLEL R FOR HIGH PERFORMANCE STATISTICAL COMPUTING, Open source software distributed from R's CRAN: <http://cran.r-project.org/src/contrib/Descriptions/RScalAPACK.html>
Impact: Distributed through more than 27 mirror sites.
5. ORNL SEMANTIC TAGGER, Open source software distributed from samatovan@ornl.gov
Impact: Used by the Department of Homeland Security (DHS) Biodefense Knowledge Center (BKC) and ADVISE program.
6. CORPUS INDEPENDENT KEYPHRASE EXTRACTION TOOL, Open source software distributed from samatovan@ornl.gov
Impact: Deployed within the DHS BKC and ADVISE systems.
7. MPIBLAST-PIO: EFFICIENT PARALLELIZATION OF NCBI BLAST, Open source software distributed from: <http://mpiblast.lanl.gov/Downloads.Stable.html>
Impact: The incorporation of performance optimizations in the NCSU-ORNL research prototype pioBLAST into the popular open-source mpiBLAST software.
8. PGRAPH: LIBRARY OF PARALLEL AND SCALABLE GRAPH ALGORITHMS, Open source software distributed from samatovan@ornl.gov
Impact: Used by a number of computational biology projects, developed in collaboration with Cray Inc.
9. PARALLEL SATLEED: LIBRARY OF PARALLEL AND SCALABLE GRAPH ALGORITHMS, Parallel Symmetrized Automated Tensor Low Energy Electron Defraction
Open source software distributed from <http://www.cs.utk.edu/~samatova/LEED/report.html>
Impact: Used by multiple scientists internationally.
10. PESS, Parallel Exact Stochastic Simulator
Open source software distributed from samatovan@ornl.gov

VI. Service to the University and Professional Societies

A. Committee Appointments

DEPARTMENT OF COMPUTER SCIENCE COMMITTEES

1. Computing Infrastructure Committee, 2007-present *New in 2008*
2. Planning for Increasing Research Funds Committee, 2007-2008 *New in 2008*
 Developed a strawman proposal on the Committee for Increasing Research Funds for the Department. The proposal has been discussed at faculty meetings and various strategies on pursuing it further are in discussion.
3. Assisting Faculty in Proposal Writing, 2007-2008 *New in 2008*
 Provide regular assistance to faculty on funding opportunities and increasing the chances for winning the proposals.
4. Peer Evaluation Committee, 2007-2008
 Conducted teaching evaluations of
 - Prof. Steffen Heber (April, 2008) *New in 2008*
 - Prof. Kemafor Anyanwu (February, 2008) *New in 2008*
5. Meeting with Faculty candidates, 2007-2008: Haixun Wang (IBM), Zhenchang Xing (U. Alberta, Canada), Ingolf Krueger (UCSD), Anotnio Roque (USC), Barbara Ryder (State Univ of New Jersey), Naren Ramakrishnan (Virginia Tech) *New in 2008*
6. Giving lectures to new graduate students in Fall, 2007 *New in 2008*
7. Advising PhD students seeking faculty positions: Harini Ramapra, Sibin Mohan, Wenhong Tian, Nirmal Desai *New in 2008*

B. National and International Activities

I have been, and continue to be, involved in numerous national activities in various roles:

- DOE/DHS/NSF WORKSHOP ORGANIZING COMMITTEES
 1. *Co-Chair*, DOE/NSF Mathematics for Analysis of Petascale Data, June 2008
 2. *Co-Chair*, DOE Genomics:GTL Program Systems Biology Network/Knowledgebase, May 2008
 3. *Co-Chair*, OE/OASCR Mathematics for Peta-scale Data Analysis & Computation, June 2007
 4. *Technical Session Chair*, DOE Workshop on Modeling and Simulation at the Exascale for Energy and the Environment, Apr, May, June 2007
 5. *Chair*, DOE SciDAC Biology Summit, August 2005
 6. *co-Chair*, DHS Threat Awareness Customer Data Integration & Dissemination, November 2005
 7. *co-Chair*, DHS Data Sciences for Homeland Security Information Management and Knowledge Discovery, September 2004
 8. *Committee Member*, DOE Data Sharing & Integration Summit, April 2006
 9. *Committee Member*, DOE SciDAC Conference, June 2005
 10. *Committee Member*, Scientific Data Management, April 2004
 11. *Committee Member*, OE Computational Infrastructure for the Genomes to Life Program, July 2002
- GRANT PROPOSAL REVIEW PANELS:
 1. DOE OBER SBIR/STTR panel, 2007
 2. DOE OASCR SBIR/STTR panel, 2007
 3. DOE OBER SBIR/STTR DE-FG01-05ER05-15 solicitation, 2004
 4. DOE OASCR SBIR/STTR DE-FG01-05ER05-28 solicitation, 2004
 5. DOE Early Career Award Panel, 2003
 6. Chair of the External Review Panel for the DOE GTL VIMSS Center, 2006
 7. DHS Internal Proposal Review BAA 06-003 Panel, 2006
 8. DHS External Proposal Review BAA 06-003 Panel, 2006
 9. DHS Pre-proposal Review BAA 06-003 Panel, 2006
 10. NSF Proposal Reviewer for the NSF05-570 solicitation, 2005
 11. DOE OASCR unsolicited proposal reviewer, 2002-present
 12. ORNL Seed Money Proposal Reviewer, 2002-present
- JOURNAL REVIEWER:
 1. BMC Bioinformatics
 2. Special Issue on Model Based Testing
 3. Journal of Bioinformatics and Computational Biology
 4. Bioinformatics
 5. ACM Computing Reviews
 6. IEEE Transaction on Pattern Analysis
- TECHNICAL PROGRAM COMMITTEE MEMBER for several conferences and workshops:

1. IEEE International Symposium on Bioinformatics and Life Science Computing (BLSC), 2007
 2. IEEE Workshop on High Performance Computing in Medicine and Biology, 2005-2006 International Conference on Scientific and Statistical Database Management (SSDBM), 2002, 2004
 3. SIAM International Conference on Data Mining, 2003
 4. IEEE Bioinformatics Conference, 2003
 5. C. Warren Neel Conference on Statistical Data Mining, 2002
 6. CLUSTER: IEEE International Conference on Cluster Computing, 2002
- SESSION CHAIR at several conferences and workshops
 - PANEL ORGANIZER at several conferences and workshops
 - BOOKS REVIEWER:
 1. Elain Rich, Automata, Computability and Complexity, Prentice Hall, 2008
 - REVIEWER for the DOE, major journals and conferences in bioinformatics and data mining

New in 2008