MS 98 and MS 111: Best Practices for Introducing Undergraduate Students to Computational and Interdisciplinary Research

Organizers: Matthias K. Gobbert and Nagaraj K. Neerchal

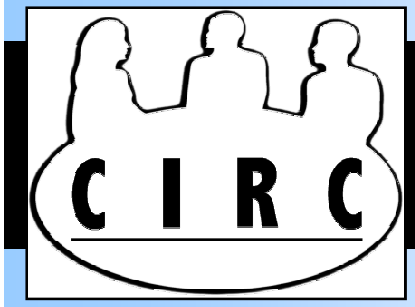
SIAM Annual Meeting 2012

MS 98:

- *Matthias K. Gobbert* and Nagaraj K. Neerchal, UMBC
- *Padmanabhan Seshaiyer*, George Mason University
- *Jennifer Pearl*, National Science Foundation
- *Jeffrey Humphreys*, Brigham Young University

MS 111:

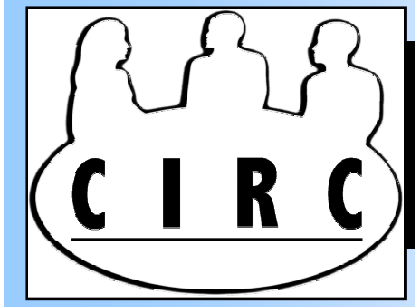
- *Peter R. Turner*, Clarkson University
- *Nabendu Pal*, University of Louisiana, Lafayette
- *Eric J. Kostelich,* Arizona State University
- *Angela B. Shiflet*, Wofford College



Profiles

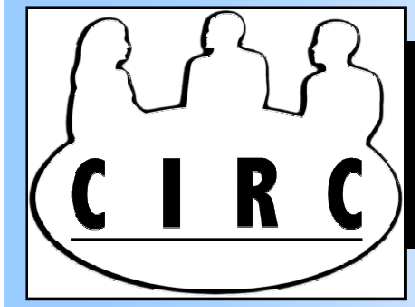
- **UMBC = University of Maryland, Baltimore County:**
  - founded in 1966 as third research university in USM;
  - 13,000 students (10,000 undergrad., 3,000 graduate);
  - 350 research faculty in 33 departments;
  - science and technology focus, particularly biology/medical research, plus Visual Arts, Public Policy, Psychology, Theatre;
  - #1 “up-and-coming” for third year in US News & World Report

- **Department of Mathematics and Statistics:**
  - 380 undergraduate majors, 50 B.A./B.S. per year;
  - M.S./Ph.D. in Applied Mathematics and in Statistics;
  - Applied Mathematics oldest graduate program at UMBC (first Ph.D. in 1975)!



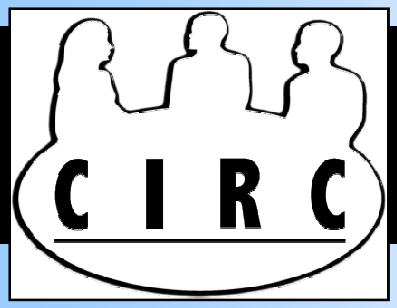
High Performance Computing Facility

- Initiated by MRI proposal in Jan. 2007 that outlined the need for **(i) hardware, (ii) sys. admin, (iii) user support, and (iv) usage policies**
- MRI proposal successful in 2008 with 23 faculty in 10 departments across campus; plus SCREMS proposal for department (4 faculty)
- **2008 ("hpc"):** 33 compute, 1 develop, and 1 user/management node; two dual-core AMD Opteron processors and 13 GB memory per node; dual-data rate (DDR) InfiniBand; 14 TB central storage
- **2009 ("tara"):** 82 compute, 2 develop, 1 user, and 1 management node; two quad-core Intel Nehalem processors and 24 GB memory per node; quad-data rate (QDR) InfiniBand; 160 TB central storage
- **HPCF user support:** since 2008 full-time RAs, www.umbc.edu/hpcf
- **Coordinated community building:** Math 627 Parallel Computing, colloquium talks in departments across campus, tech. rep. server, meetings with administrators, follow-up grant proposals, etc.



Center for Interdisciplinary Research and Consulting (CIRC)

- Nagaraj K. Neerchal, Statistics, and Matthias K. Gobbert, Mathematics
- Started in 2003 in form of regular class with client-based projects; some results of these: journal paper (Soane) and long-term RA position (Sun)
- **CIRC makes department's expertise in statistics and in applied mathematics available to the community on campus and beyond in consulting format**
- Mathematics and statistics students gain hands on interdisciplinary experience vital for industry and academia jobs = industrial mathematics
- CIRC has full-time RAs since 2005, www.umbc.edu/circ
- **Benefits to students:** experience, publications, presentations, RAs
- **Benefits to department:** visibility, connections, money, uniqueness, synergy between programs, energy level!



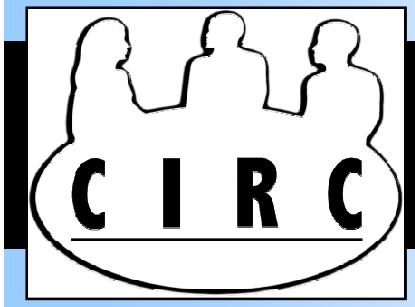
REU Site: Interdisciplinary Program in High Performance Computing

Department of Mathematics and Statistics
University of Maryland, Baltimore County
www.umbc.edu/hpcreu
hpcreu@umbc.edu

Undergraduate Research on the Fast Track: From Nothing to Publication in Eight Weeks

Directors of the REU Site:

Nagaraj K. Neerchal and Matthias K. Gobbert



Concept of REU Site

8 weeks, team-based with 4 undergraduate students, 3 (over-lapping) phases:

- Phase I – 2 weeks: 3-credit transferrable course on scientific, statistical, and parallel computing, introducing Linux, C, MPI, Matlab, R
- Phase II – 5 weeks: research on application project (with computational focus) from outside mathematics/statistics
- Phase III – 1 week: complete documentation of work in form of HPCF tech. rep., poster, talk, webpage
- Accompanying professional development program

Shown from perspective of undergraduates in following!

Phase I – Weeks 1 to 5

Training in scientific, statistical, and parallel computing:

- Transferrable three-credit course
- Introduction to Linux, C, MPI, Matlab, R
- Lectures complemented by computer labs with graduate Tas
- Homework done by assigned teams

***Simultaneously*, several potential clients present their projects:**

Phase II – Weeks 1 to 7

Research on application problem in team of 4 undergraduates:

- Each team has faculty mentor and dedicated graduate RA
- Team members know each others' strengths and preferences by now!
- Updates to client typically once a week, in person, by conference call, or similar

Phase III – Weeks 3 to 8

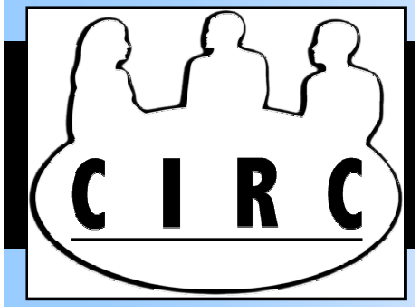
Complete range of documentation of results:

- Deliverable to client can be computation, data analysis, code, visualization, webpage, advice, or others
- Tech. rep. posted on HPCF webpage, other publication considered
- Presentations in poster form and for oral presentation at the UMBC Summer Undergraduate Research Fest (SURF)
- Project webpage at REU Site www.umbc.edu/hpcreu

Professional Development Program

Obvious parts:

- Introduction to LaTeX, preparation of poster, talk, webpage
- GRE preparation course
- Presentations by Dean of the Graduate School on graduate school application; other presentations, e.g., on posters by Assistant College Dean and on career choices by Director of Academic Advising



Professional Development Program

Many not-so-obvious parts:

- Tech. report HPCF-2012-X posted on HPCF Publications webpage
- 'Vertically Integrated' support for each team with graduate RA and faculty; local and/or returning students as peer mentors
- VIP visits by President, Provost, Dean, for instance, and also GPD and editor of *UMBC Review: Journal of Undergraduate Research and Creative Works* as example of undergraduate journal (e.g., SIURO)
- Interview all visitors about their career; each team gives 'elevator speech' to visitors; students write report about visit.
- Make *explicit* the guidance on research techniques, including tracking sources, documenting, issues of integrity, etc.
- Share our experiences for graduate school advice, for instance, share our perspective on admissions
- Field trips, e.g., to NSA, NASA, as well as to D.C. and Baltimore

Projects 2010 and 2011

- **Enabling Physiologically Representative Simulations of Pancreatic Beta Cells**

  Clients: Bradford Peercy, Math & Stat, UMBC, and Arthur Sherman, NIH
- **Parallelization of Matrix Factorization for Recommender Systems**

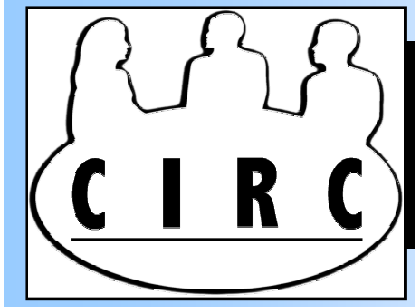
  Client: Robert Bell, AT&T Labs, Florham Park, NJ
- **Assessment of Simple and Alternative Bayesian Ranking Methods Utilizing Parallel Computing**

  Client: Martin Klein, U.S. Census Bureau
- **Sampling Within k-Means Algorithm to Cluster Large Datasets**

  Client: George Ostrouchov, Oak Ridge National Laboratory, Oak Ridge, TN
- **Optimization of Computations Used in Information Theory Applied to Base Pair Analysis**

  Client: Patrick O'Neill and Ivan Erill, Biological Sciences, UMBC
- **Intel Concurrent Collections as a Method for Parallel Programming**

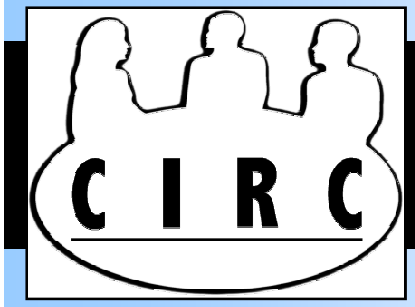
  Client: Loring Craymer, DoD Center for Exceptional Computing



REU Site: Interdisciplinary Program in High Performance Computing

Vital stats now:

- **Funded by NSF** for 8 students in Summers 2010 and 2011, additional students via UMBC Meyerhoff / NSA funded program; renewed for 12 students in Summers 2012, 2013, 2014 **funded jointly by NSF and NSA**
- **In 2012**: 3 teams of 4 students, 2 graduate TAs, 3 graduate RAs, 2 peer mentors, some of these leveraging joint funding by HPCF, CIRC, UMBC, and the department
- **Projects in 2012** with clients from Sandia National Lab, UMBC Office of Institutional Research, and Maryland Department of Natural Resources



Lessons Learned

- **Track from the start:** status of applicants (gender, class standing, race/ethnicity, disability, veteran, and a lot more on application form) ; examples 2010: 2 African-American, 2 Hispanics; 2011: 4 African-American, 1 Asian, 1 Hispanic, 1 veteran; 2012: 1 Asian, 2 Hispanic, 1 veteran
- **Document and present:** schedule and more details on webpage, take photos (people and events), collect reports of all events including training
- **Team work** as goal in itself, then manage pro-actively and explicitly
- **Use modern technology and methods:** example iPad
- **Stay in touch** for longitudinal tracking and for documenting 'future' outcomes such as presentations at home institutions and conferences

For all details on our program: www.umbc.edu/hpcreu

Acknowledgments: NSF-REU, NSF (MRI, SCREMS), UMBC, HPCF, CIRC