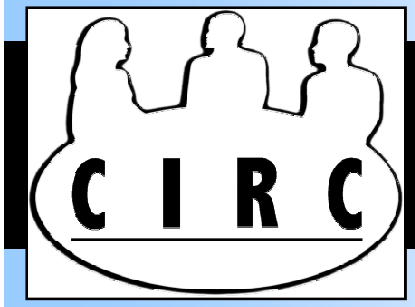
Matthias K. Gobbert, Nagaraj K. Neerchal, Padmanabhan Seshaiyer

Special Session on Undergraduate Research and its Impact on Students and Faculty

Matthias K. Gobbert and Nagaraj K. Neerchal
Undergraduate Research on the Fast Track: From Nothing to Publication in Eight Weeks

Interdisciplinary Program in High Performance Computing
Department of Mathematics and Statistics
University of Maryland, Baltimore County (UMBC)
www.umbc.edu/hpcreu

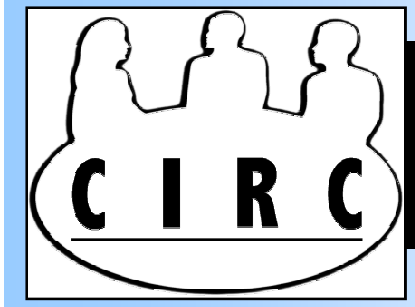
Example Project: Gemma Gearhart, Shuai Jiang, Thomas J. May, Jane Pan, RA Samuel Khuvis, Mentor Matthias K. Gobbert, Clients Bradford E. Peercy and Arthur Sherman, *Dynamics of Computational Islet Simulations*



Profiles

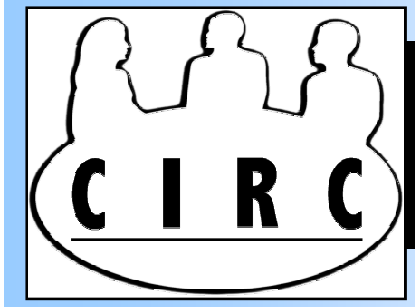
- **UMBC = University of Maryland, Baltimore County:**
  - founded in 1966 as third research university in USM;
  - 14,000 students (10,000 undergrad., 3,000 graduate);
  - 350 research faculty in 31 departments;
  - science and technology focus, particularly biology/medical research, plus Visual Arts, Public Policy, Psychology, Theatre;
  - #1 “up-and-coming” in US News & World Report since 2009, President Freeman A. Hrabowski on TIME Magazine Top 100 List

- **Department of Mathematics and Statistics:**
  - 385 undergraduate majors, 55 B.A./B.S. Math/Stat per year;
  - M.S./Ph.D. in Applied Mathematics and in Statistics;
  - Applied Mathematics oldest graduate program at UMBC (first Ph.D. in 1975)!



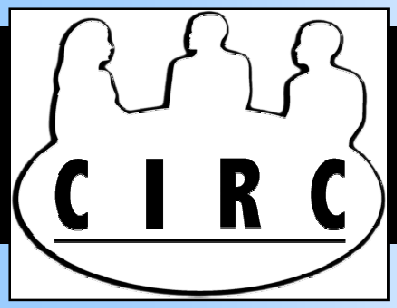
Center for Interdisciplinary Research and Consulting (CIRC)

- 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 and long-term RA position
- **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!



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**
- **2009:** MRI grant to 23 faculty in 10 departments (2008); 86 nodes; two quad-core Intel processors and 24 GB memory per node; quad-data rate (QDR) InfiniBand; 160 TB central storage
- **2013:** MRI grant to 30 faculty (2012); extension by 72 nodes with two eight-core Intel CPU and 64 GB memory per node, including 19 hybrid nodes with two NVIDIA K20 GPU and 19 hybrid nodes with two 60-core Intel Phi accelerators; extension of network; extension of storage; gift from NASA (first large-scale computer gift to UMBC!)
- **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.



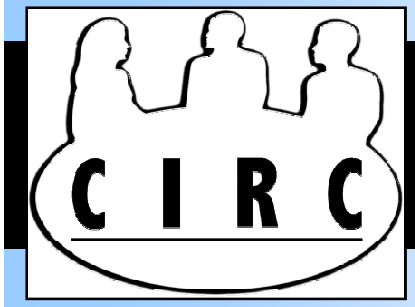
REU Site: Interdisciplinary Program in High Performance Computing

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

Undergraduate Research in an REU Site: 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., SURF poster, talk, webpage
- Accompanying professional development program

Shown from perspective of students in following!

Phase I – Weeks 1 to 3

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

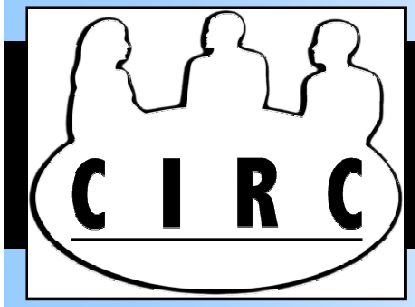
Complete range of documentation of results:

- 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 (in collaboration with other UG programs!)
- 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-201X-Y 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* 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, NIH, as well as to D.C. and Baltimore

Projects in 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**

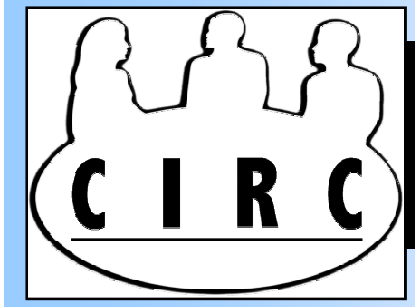
  Client: Robert Bell, AT&T Labs, Florham Park, NJ
- **Assessment of Simple and Alternative Bayesian Ranking Methods Utilizing Parallel Computing**

  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

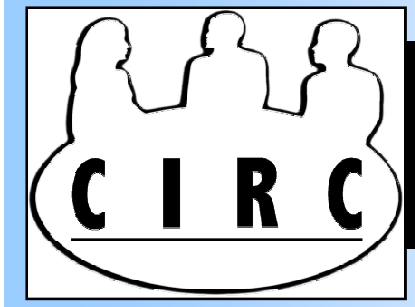


Projects in 2012 and 2013

- **Graph 500 Performance on a Distributed-Memory Cluster**
  Clients: Richard C. Murphy, Sandia National Lab, and David J. Mountain, NSA
- **Water Quality Monitoring of Maryland's Tidal Waterways**
  Client: Brian R. Smith, Maryland Department of Natural Resources
- **Simulation of a University as a Dynamical System**
  Client: Michael Dillon, Office of Institutional Research, UMBC
- **Block Cyclic Distribution of Data in pbdR and its Effects on Efficiency**
  Client: George Ostrouchov, Oak Ridge National Laboratory
- **Identifying Nonlinear Correlations in High Dimensional Data with Application to Protein Molecular Dynamics Simulations**
  Client: Ian F. Thorpe, Department of Chemistry, UMBC
- **Contention of Communications in Switched Networks and Clustering of Multidimensional Data Sets**
  Clients: Philip J. Farabaugh, Department of Biology, UMBC, and Christopher Mineo and David J. Mountain, NSA
- **Investigating Oscillation Loss in Computational Islets**
  Clients: Bradford E. Peercy, UMBC, and Arthur Sherman, NIH

Graph 500 Performance on a Distributed-Memory Cluster

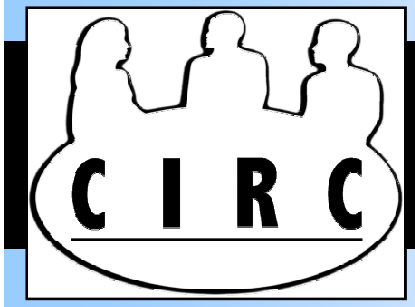
- **Benchmark quantifies memory access speed of computer.** The team implemented the Graph 500 benchmark on our cluster tara and submitted the results for the November 2012 ranking
- **Trip to conference Supercomputing 2012 in Salt Lake City** for the formal announcement of the ranking as #98:
- **First reunion of a team** led to travel funding
- **SIAM Annual/CSE mtg.s, AMS Sectional Meeting** Special Session on Undergraduate Research



REU Site: Interdisciplinary Program in High Performance Computing

Vital stats:

- **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**
- 2010: 8 students in 2 teams, 2 graduate TAs, 1 graduate RA
- 2011: 13 students in 4 teams, 2 graduate TAs, 3 graduate RAs, 2 peer mentors, 1 additional faculty
- 2012: 12 students in 3 teams, 3 graduate TAs, 3 graduate RAs, 2 peer mentors
- 2013: 16 students in 4 teams, 2 graduate TAs, 2 graduate Ras, 1 additional faculty
- 2014: 12 funded students, at least 1 self-funded, local team, at least 2 graduate TAs/RAs, 2 additional faculty
- Staffing beyond the 2 co-directors leverages joint funding by HPCF, CIRC, UMBC, and the department



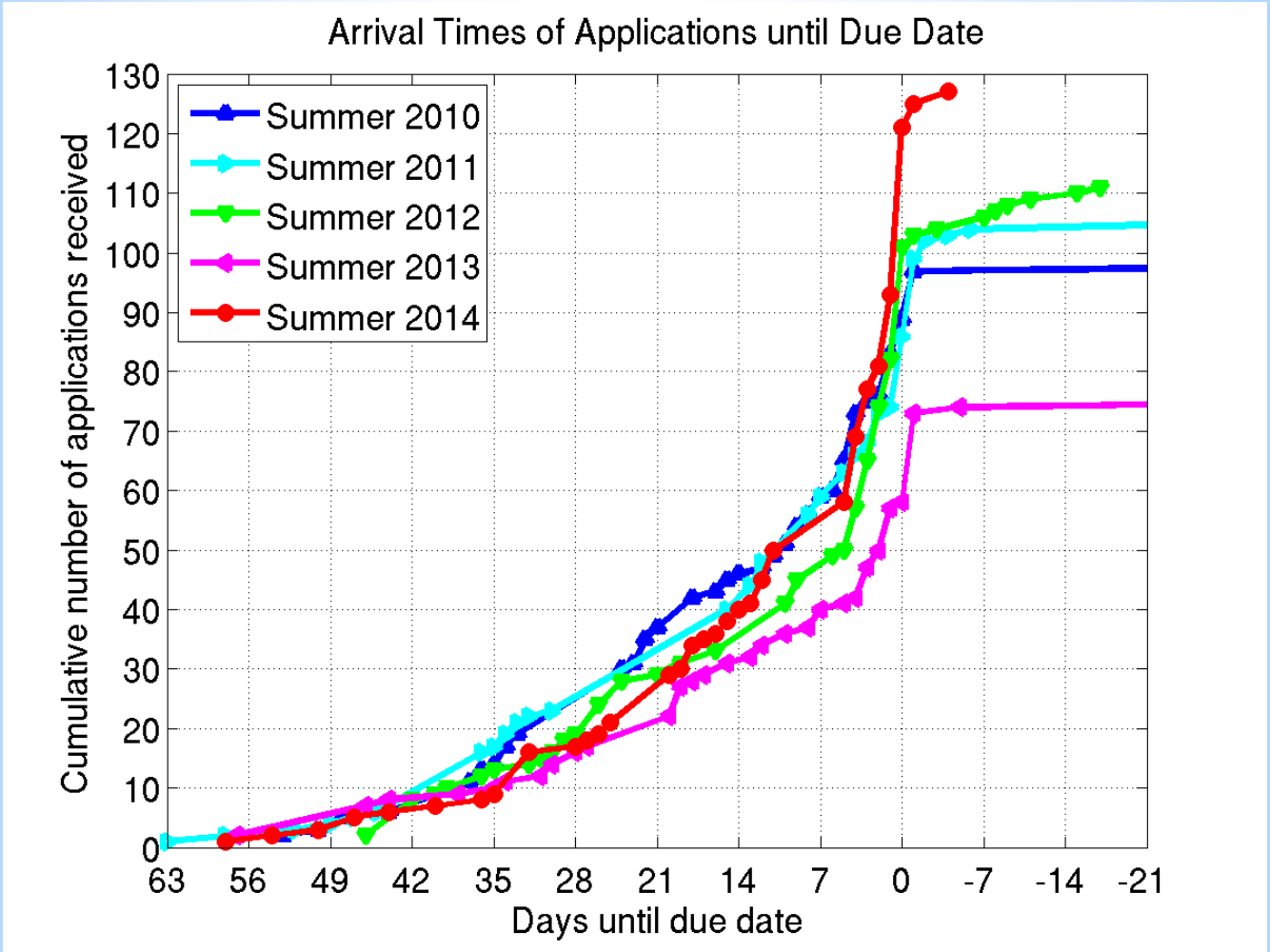
Lessons Learned

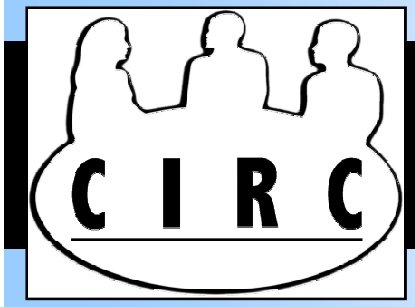
- **Track from the start:** status of applicants (gender, class standing, race/ethnicity, disability, veteran, and a lot more on application form) ; examples 2010: 2 African-American, 2 Hispanics; 2011: 4 African-American, 1 Asian, 1 Hispanic, 1 veteran; 2012: 1 Asian, 2 Hispanic, 1 veteran; 2013: 4 Asians, 1 Hispanic, 4 African-American; 2014: 3 Asians, 2 Hispanic, 2 African-American, 1 veteran, 1 disabled.
- **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:** examples Blackboard, 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, NSA, UMBC, HPCF, CIRC

Application Arrivals 2010-2014





Example Project from 2013

- **Team members:** Gemma Gearhart, Shuai Jiang, Thomas J. May, Jane Pan (UMBC Meyerhoff Scholar, NSA grant to Meyerhoff Program)**,** **Graduate RA:** Samuel Khuvis, **Faculty mentor:** Matthias K. Gobbert, **Clients:** Bradford E. Peercy (UMBC) and Arthur Sherman (NIH)
- **Project title: Dynamics of Computational Islet Simulations**
- **Presentations:** 1 of 6 oral presentations at SURF, poster at SURF, Gobbert at BEER (International Symposium on **B**iomathematics and **E**cology **E**ducation and **R**esearch)
- **Publications:**
  - Jane Pan, Gemma Gearhart, Shuai Jiang, Thomas J. May, *UMBC Review: Journal of Undergraduate Research*, 2014 (to appear at URCAD on April 23)
  - Gemma Gearhart, Shuai Jiang, Thomas J. May, Jane Pan, Samuel Khuvis, Matthias K. Gobbert, Bradford E. Peercy, and Arthur Sherman, *Proceedings of BEER 2013*, 2014 (accepted)