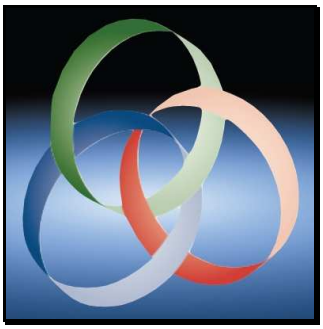
High-Performance Research Computing and Cyberinfrastructure:

An Interim Assessment of the Demand-Side at the University of Michigan

A Report to the Community from the
Office of the Associate Vice President for
Research Cyberinfrastructure

Daniel E. Atkins
Katherine A. Lawrence
Max Bronstein



May 2009

Table of Contents

Background and Activities of the Office of the Associate Vice President for Research Cyberinfrastructure (AVPR-CI) 2

Community Survey 3

Town Hall Meetings 7

Appendix 1: OVPR-CI Questionnaire 9

Appendix 2: Key Points from Town Hall Meetings 11

March 26 Meeting on Central Campus 11

March 31 Meeting on North Campus 12

Background and Activities of the Office of the Associate Vice President for Research Cyberinfrastructure (AVPR-CI)

On September 1, 2008, Professor Daniel Atkins began his appointment as the newly established Associate Vice President for Research, Cyberinfrastructure (AVPR-CI) within the Office of the Vice President for Research (OVPR). The office of the AVPR-CI has defined its mission very broadly as follows:

> *Our mission is to catalyze partnerships to provision, to apply in transformative ways, and to continuously improve advanced cyberinfrastructure at the University of Michigan. Although our primary focus is on research, we also aspire to empower learning and societal engagement at a higher level of excellence, positive impact, scale, and global reach.*

Achieving this mission requires on-going communication, participatory design, and alignment of mutual self interest between those who provision cyberinfrastructure (supply side), those who apply cyberinfrastructure to discovery and learning (demand side), and those who produce research results that are relevant to improving the cyberinfrastructure environment in the future (both new supply and new demand)—all with associated educational requirements. To that end, Atkins recruited several staff to assist in facilitating the growth of awareness of and resources for research computing at UM. This group—the OVPR-CI—began several initiatives during the 2008–2009 academic year in order to gain a better understanding of UM's use of research computing.

This report details what we have learned so far about the demand side for cyberinfrastructure at UM through a community survey and town hall meetings. The survey of the UM community used a web-based questionnaire that was distributed with the assistance of the Associate Deans for Research, Rackham Graduate School, and a selection of relevant e-mail groups. In March 2009, OVPR-CI hosted two town hall meetings. These meetings provided the community with an opportunity to voice their concerns and discuss key issues and challenges regarding research computing at UM. These meetings were promoted in conjunction with the survey.

In addition to the survey and town halls, the OVPR-CI has pursued the following methods of gathering input from and sharing knowledge with the community:

- Meeting with Deans, Associate Deans for Research, and senior administrators.
- Meeting with faculty and other stakeholders who have a keen interest in research computing and the enabling cyberinfrastructure.
- Finding and linking with "research computing committees" if they exist within schools and colleges.
- Creating a website within the OVPR domain to help share information and build community (http://www.research.umich.edu/ci/). This site provides an overview of OVPR-CI activities and links to internal and external computational research.
- Establishing a moderated listserv to spread news of interest to the community. (To join the ResearchComputingNet listserv, go to https://listserver.itd.umich.edu/cgi-bin/lyris.pl?enter=researchcomputingnet)
- Creating a registry that will allow people within UM to search and identify others who share their interests (or have complementary skills) with regard to research computing.

- Supporting the formation of a Virtual School for Computational Science.
- Identifying individuals who may be willing to play leadership roles in future activities or initiatives.

Community Survey

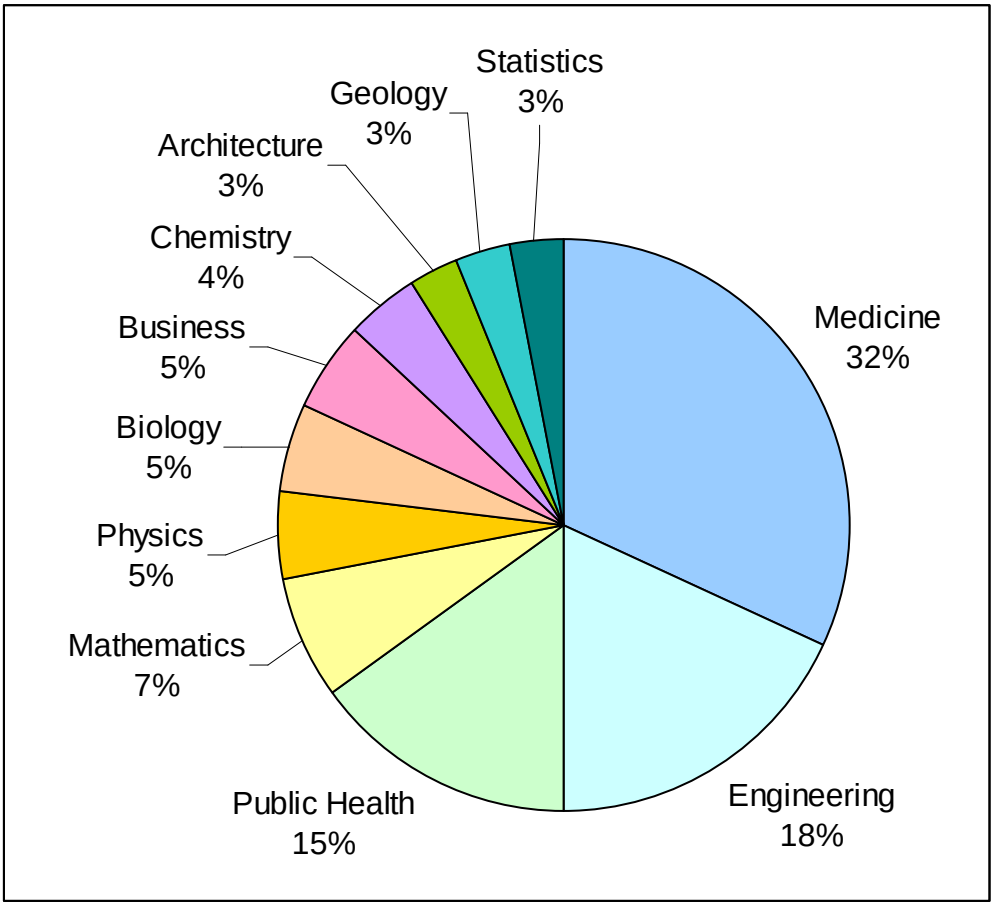
In March 2009, the OVPR-CI broadly distributed a questionnaire to UM faculty, staff, and students in order to identify individuals interested in or engaged with high-performance research computing. The goals of the survey were to:

- Identify interested members of the community so as to establish a registry that would enable connections and collaborations across the campus.
- Gather information about research interests and resources used for high-performance computing (HPC) for research at UM.
- Understand community members' primary concerns about high-performance computational resources at UM.
- Establish a base of community members who would be available to provide further input or feedback about CI initiatives.
- Identify the viability of tying into existing social network services as a way to build the research computing community at UM.

Responses to the open-ended questions about research interests and areas of concern were coded and grouped, and the top responses are described in the remainder of this section. The degree of elaboration provided by respondents varied, so this analysis required some degree of extrapolation to appropriately group responses. (See Appendix 1 for the full list of survey questions.)

Of the 278 responses, 47% were faculty or research scientists, 18% staff, and 35% graduate students. Figure 1 identifies the primary disciplinary areas represented in the responses to the questionnaire. Almost 1/3 of respondents were from the Medical School. Other top disciplines were engineering (18%) and public health (15%). The remainder of responses represented departments in LS&A and professional schools.

Figure 1: Primary Disciplinary Areas Represented in Survey Responses (n=220)



A key aim of the survey was to identify areas of concern relating to high-performance computational resources (e.g., access, data storage, visualization, support, training, collaboration, etc.) at UM. Figure 2 displays the top 11 issues. By far the most prominent issue is access to HPC (39% of respondents indicated that this is an issue). Approximately a quarter of respondents are concerned with support, storage, or training. Additional key areas of concern include collaboration, faster CPUs (central processing units used for computation), security, visualization, centralization, greater memory (for performing calculations), and backup.

Although some respondents simply listed "access" as a concern without providing further detail, many respondents provided a more thorough commentary. For example, one staff member elaborated:

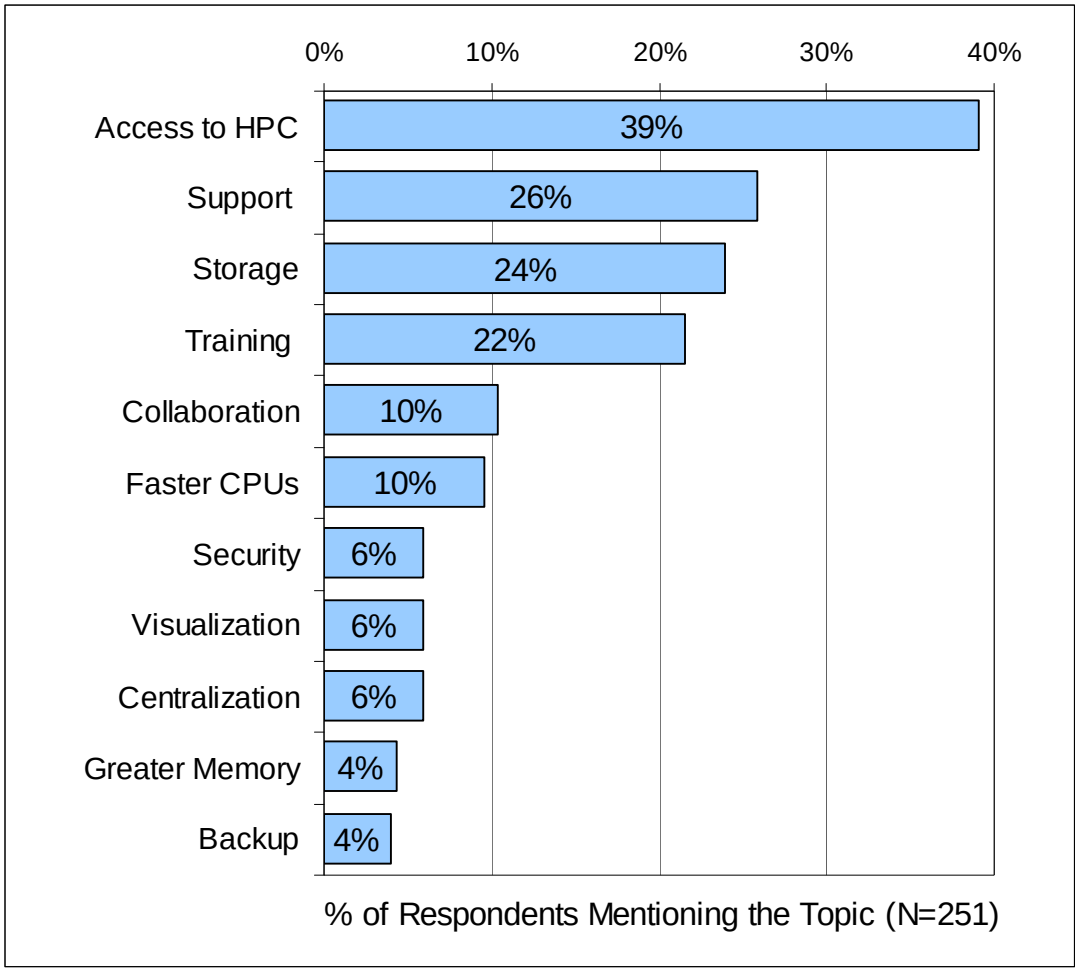
- "**Ease of access**, that is, **low barriers to use the service**, obtaining an account, low fees if any, etc., particularly for students and junior faculty." (staff member)

Other frequently occurring concerns related to adequate resources for storing data for both short-term (backup) and long-term (archiving) needs. Respondents also mentioned that support for HPC resources was essential to achieving their research goals. Some respondents raised multiple issues:

- "I'm concerned with having **easy access to clusters** without having to wait for computer time. I'm also concerned with having **adequate data backup** for my lab without having to set up my own server. I'm also concerned with getting the **basic computing support** I need for desktop computers in my lab (Macs, whereas the department is more PC focused), as well as getting help with setting up a small cluster for my lab." (faculty)

- "We need high capacity **low cost network accessible scratch disks**. Right now the UM community has several high-quality storage options, all at exorbitant cost. We need ultra low cost, super high capacity (say a single user could gain access to 10 TB) but no overhead costs of nightly backup, or redundancy, or disaster recovery. Just cheap, copious storage, for huge test or reproducible data sets, available on the network." (graduate student)

Figure 2: Primary Concerns about High-Performance Computing (HPC)

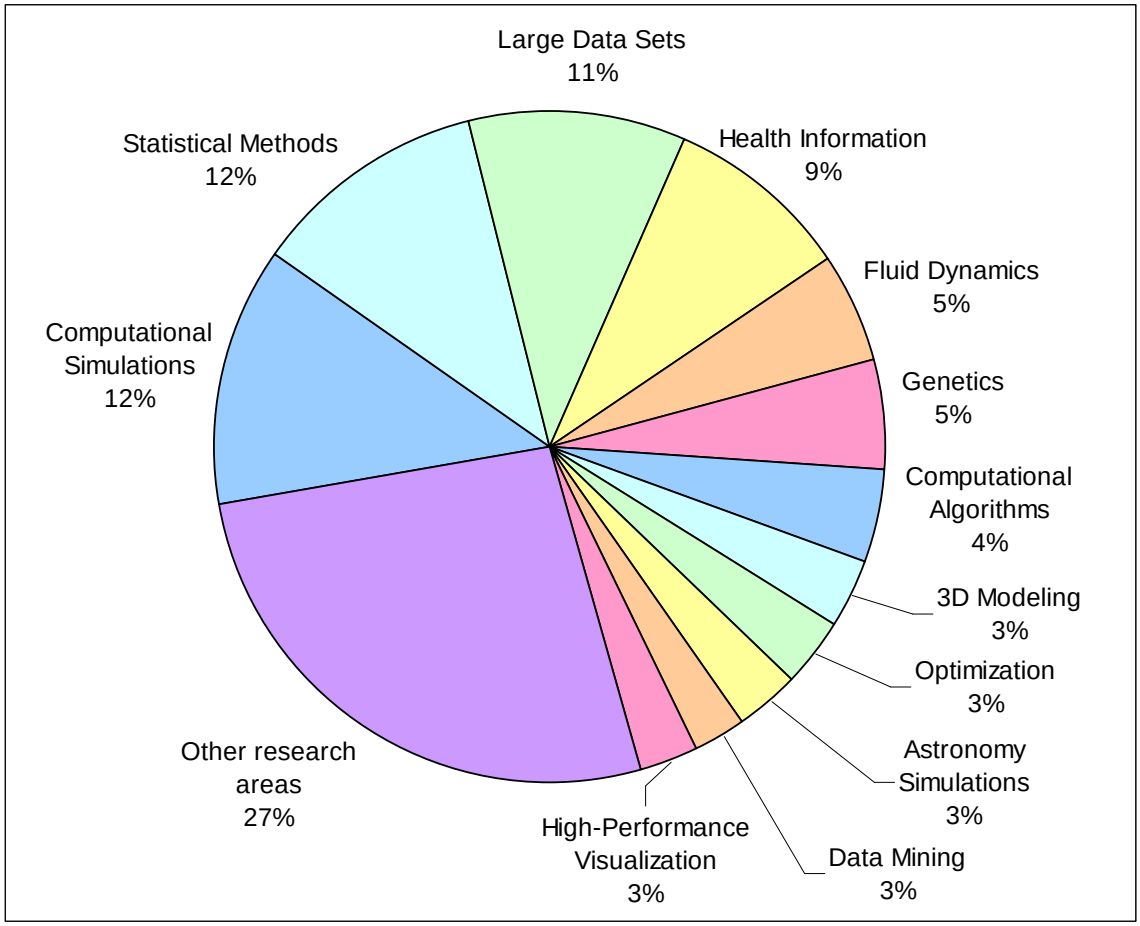


In addition, several respondents pointed out that UM's current approach to computing is highly decentralized and inefficient. Many respondents also called for opportunities for collaboration among researchers:

- "**No centralization of resources**, and a general lack of high speed gigabit (or faster) connections between major infrastructure. **HPC is much like electricity and water to research nowadays, and ideally needs to be provided to all researchers as needed.**" (staff)

- "I think this is a great program to investigate how to use resources to better fit the needs of research here at U of M. Despite modeling very different types of problems, the resources and tools used are often the same across many disciplines in scientific computing and **collaboration and feedback** from other areas on the **scientific computing aspect would be helpful for all involved.**" (graduate student)

Figure 3: Major Areas of Research Using High-Performance Computing
(n=347; some respondents listed more than one area of research)



Lastly, participants were asked to describe the nature of their computational research (see Figure 3). As with primary concerns, respondents provided varying amounts of detail in their remarks. Many researchers use HPC to conduct various types of computational simulations (12%) and statistical analyses (e.g., Monte Carlo or Bayesian; 12%). Respondents also mentioned working

with very large data sets to accomplish their research goals (11%). Given that many of the respondents were from the medical school, a substantial proportion of researchers conduct computational research that contains health information (9%).

These survey responses provided OVPR-CI with important insight into the areas of computing that need special attention. In addition, these responses will allow the OVPR-CI to assemble a registry database based upon common interests, concerns, and academic units. This will allow researchers to identify and contact potential colleagues and mentors at the University, and it will support targeted communications to be sent to these cohorts.

Town Hall Meetings

In late March 2009, the OVPR-CI held two town hall style meetings in order to engage and obtain feedback from the UM community. Approximately 126 faculty, staff, and students actively participated in these meetings. At each meeting, Associate Vice President Atkins presented an overview of cyberinfrastructure, e-science, and outlined issues facing computational researchers at the University of Michigan based on the responses to the survey. Participants were then prompted to discuss and respond to the following questions:

1) What type of organization and resources do we need to facilitate the broader impact (transformative) and application of cyberinfrastructure for research and learning?

2) Why should UM increase internal investment in HPC resources? What would be the major benefits of a more robust computing environment?

These are some of the major issues and recommendations raised by the participants. Appendix 1 provides a more detailed summary.

- UM would benefit from a centralized, University-wide HPC resource. A centralized system would enhance efficiency, provide researchers with increased computational capacity, and should be available to all faculty, especially early career investigators.

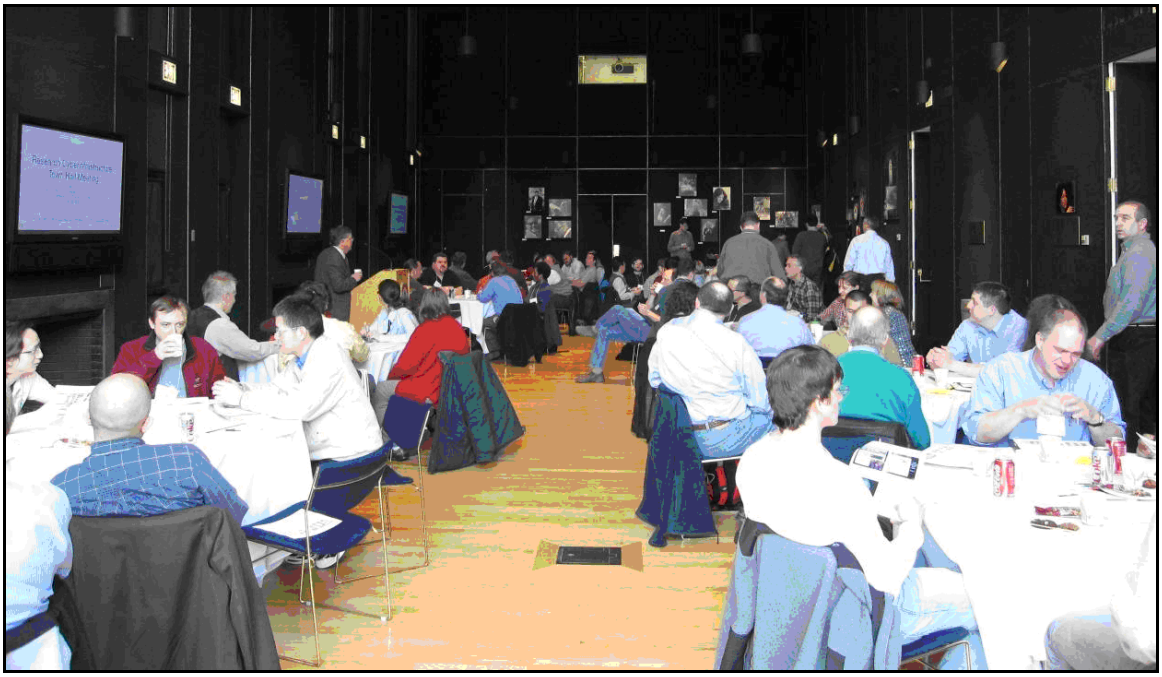
- Without a robust HPC environment, the University is at risk of falling behind competing institutions. This results in difficulty in recruiting and retaining both students and faculty with interests in computational research. In addition, competition for external grants and other resources is increasingly global and selective; a greater investment in HPC could add value to these pursuits.

- Faculty need dedicated, highly trained support staff to allow easy access and maintenance of HPC. Training could also be offered to faculty or staff who are interested in learning more about computational research.

- Currently, there is greater demand for the lower end of high performance computational power. However, these lower end systems could serve as a stepping stone to greater use of high-end grid computing such as the TeraGrid. Purchasing a high-end system for UM could raise the visibility of the University and enhance competitive standing.

- Collaboration should be fostered between both faculty and support staff in order to advance research and develop best practices.

Figure 4: March 26 Town Hall on Central Campus



These meetings provided the computational research community with an opportunity to gather and discuss common issues and potential solutions. Participants also had the opportunity to network with other computational researchers. The OVPR-CI is incorporating these valuable comments and recommendations into initiatives that will better serve the computational research community.

Appendix 1: OVPR-CI Questionnaire

Provide input to UM high-performance research computing initiatives

This very short survey is sponsored by the UM Office of the Associate Vice President for Research Cyberinfrastructure.

High-performance computing (HPC) for research includes computational modeling, simulation, and prediction; data management and data mining; visualization; computer-mediated collaboration; and associated consulting and training services.

Our goals are:

- **to nurture an active and collaborative community among researchers who use HPC**
- **to invest significant resources into the enabling cyberinfrastructure**

Please help us understand *your* needs by taking a few minutes to answer the following questions. By identifying those interested in participating in this initiative, we will create a community registry, but we will not share your information without your consent.

Name:
Uniqname:
Department:

What is your position at the University?

- ○ Faculty/Research Scientist
- ○ Staff
- ○ Post-Doctoral Fellow
- ○ Graduate Student
- ○ Undergraduate Student
- ○ Other (please specify)

Describe briefly (in a few sentences or phrases) *your research* using high-performance computing resources, including any websites or blogs that provide additional detail.
For example: specific domain, research topics, methods, special expertise, etc.

Describe briefly the *resources* you use for your high-performance computing research. For example: TeraGrid, specific computing centers or labs, local clusters, etc.

May we include the above information in a directory for the UM computational research community? Such a directory would help community members locate people with specific expertise.

- ○ No
- ○ Yes
- ○ Yes, with the exception of … (please specify)

In which of the following activities would you be willing to participate?
Check all that apply.

- ☐ Town hall (informal opportunity to learn more and voice your opinions)
- ☐ Interview (one-on-one, scheduled at your convenience)
- ☐ Additional survey (data gathering of more specific technology and support requirements)
- ☐ Focus groups (small group meetings, possibly over lunch)
- ☐ None of the above

Please provide names and/or email addresses of other members of the UM community (staff, students, post-docs, faculty) who you believe we should especially encourage to participate in this initiative.

Are you a member of any of the following social/professional networks?
Check all that apply.

- ☐ LinkedIn
- ☐ Facebook
- ☐ Plaxo
- ☐ Friendster
- ☐ LiveJournal
- ☐ Orkut
- ☐ MySpace
- ☐ ResearchGATE
- ☐ scispace.net
- ☐ 2Collab
- ☐ Badoo
- ☐ Other (please specify)

What are your primary concerns with regard to high-performance computational resources at UM? For example: access, data storage, visualization, support, training, collaboration, etc.

Additional comments:

Appendix 2: Key Points from Town Hall Meetings

March 26 Meeting on Central Campus

Faculty Recruitment and Competitiveness

- Lacking a robust HPC environment makes faculty recruiting more difficult, especially for computationally intensive research.
- Competition for grants and faculty is truly global and increasingly selective, we need greater HPC investment to support these ends.

Centralized, University-Wide, HPC Resource

- Centralized system will be much more efficient, will provide a singular resource for researchers.
- One concern is that some could exploit the queuing system and this will require robust protocols and administrator oversight.
- Ensure that all faculty have access (especially early career investigators); senior faculty should not get priority.

Support, Training, and Access to Expertise

- Researchers need access to highly trained staff with expertise in HPC; hardware is only as good as those who manage it.
- Researchers could benefit from training courses designed to give better understanding of HPC hardware, software, and applications.

Collaboration

- Collaborative efforts must be fostered and rewarded; interdisciplinary efforts are essential to faculty learning and development.
- Need increased collaboration among staff as well, which could help them develop usage policies and best practices.

Hardware

- There is substantial demand for lower-end needs in the computational research community, but investing in a major high-end system could increase the visibility of UM and serve as a stepping-stone to TeraGrid usage.
- Provide training clusters for optimizing code and providing experience to new computational researchers or support staff.
- Should UM be investing in cloud computing?
- UM needs to raise visibility of the HPC resources available to faculty.

Fostering a Culture

- Involve students in this HPC initiative—both undergrads and graduate students.
- Create more avenues for educating students about computation, algorithms, programming languages. A Rackham certificate program is available and is being revised in College of Engineering.
- Host an HPC symposium.

March 31 Meeting on North Campus

Faculty Recruitment and Competitiveness

- UM is falling behind competing institutions and we're losing students and faculty to them.

Centralized, University-Wide, HPC Resource

- Ensure that participants who contribute to a centralized resource get more out than they put in.

Collaboration

- The community needs enhanced interaction and visibility.
- Support staff can help facilitate collaboration.

Hardware

- Having a top-10 machine would gain national attention.
- Leverage local resources for testing code, then migrate to larger resources.
- Ensure adequate data storage for long term and short term needs.
- Encourage PIs to write HPC into proposals.

Fostering a Culture

- Measure HPC impact on UM.
- Create a training program in Computational Science. Grid networking should be part of computational science education.
- Sponsor a CAC annual forum or symposium.
- Provide students with early exposure to HPC and grid computing.
- Create a [more visible] certificate program through Rackham.
- Ensure transparency in building a culture and facility.

Support, Training, and Access to Expertise

- Ensure low threshold for researchers to use services.
- Faculty time is limited and needs to be leveraged efficiently. A key incentive for encouraging faculty participation in centralized services is to demonstrate that it can leverage their time.
- Need staff to support users in creating parallel, scalable code.