

STRATEGIC VISION FOR THE MINNESOTA SUPERCOMPUTING INSTITUTE

*Long-Range Planning Committees for Computation in Engineering, Math
and Physical Sciences and for Computation in Biosciences*

April 27, 2007

The Minnesota Supercomputing Institute plays a pivotal role in the University of Minnesota's campaign to become one of the top-three public research institutions. The high-performance computing needs of the University's research community will be served best through a centralized facility with associated satellite laboratories. Several recommendations are made for administrative and programmatic changes to improve the University's high-performance computing infrastructure and to achieve its full potential for advancing the University's strategic goals.

As a steward of excellence in STEM (science, technology, engineering, and mathematics) and bio-medical discovery, the University of Minnesota has for more than 30 years embraced scientific computing as an indispensable component of its research endeavors. The University's past investments in high-performance computing (HPC) have paid off handsomely. However, while the role of high-end computational research has dramatically increased world-wide over the past 15 years (e.g., the Department of Energy's Accelerated Strategic Computing Initiative or Japan's Earth Simulator), the University's investment in HPC has fallen, and the University of Minnesota must reclaim its leadership position in this field. To this extent, two long-range planning committees (see Appendix A for a list of committee members) have assessed the current status of the Minnesota Supercomputing Institute (MSI) and developed a strategic vision for its future.

Continued and enhanced support for a center of excellence in high-performance computing research and support services is essential for the University's strategic goals because:

1. *HPC is an indispensable component of cutting-edge research.* It is now well established that HPC jointly with experimentation and theory are the foundation of scientific discovery. Without HPC support, exciting and important research areas that the University of Minnesota aspires to play a significant role, such as renewable biofuels, biomedicine, and nanotechnology, cannot be explored to their full potential. HPC is enabling interdisciplinary and transformative research that is vital for a world-class research institution. (See Appendix B.)
2. *MSI provides a direct benefit to the University in terms of research productivity, recognition, and federal funding, all measures of success for a public research institution.* Since 1998, researchers have published more than 2000 scientific papers that acknowledged MSI support. In 2006, research grants totaling about \$90M include HPC in the work plan. External funding and peer-reviewed publications are major gauges of productivity for academic institutions and the amount and quality of work facilitated by MSI are simply unsurpassed.
3. *MSI provides a direct benefit to the University in terms of the ability to attract and retain high-quality faculty.* MSI has played a pivotal role in recruiting and retaining several of our most distinguished faculty members. More than 300 principal investigators at the University benefit from MSI resources. In a survey commissioned by the two committees, the user community has indeed responded in overwhelmingly positive terms with respect to the impact MSI has on their research. (See Appendix C.)
4. *Centralized resources are the most cost effective way to support HPC at the University.* MSI provides the critical mass that is crucial for a wide spectrum of research efforts, from quantum chemistry to population genetics. Exceptions to the effective employment of centralized resources are fields for which data security is paramount, such as clinical informatics. Distributed resources would result in increased costs for hardware, software, maintenance, security, and support. Without MSI, start-up packages for researchers relying on high-

performance computers would increase substantially, as well as requirements for space and support personnel.

5. The major federal funding agencies (NIH, NSF, DOE, DOD) are formulating strategic visions for STEM and bio-medical discoveries that are inherently based on computational sciences and large, centralized computing resources. From NSF's Cyberinfrastructure and NIH's Bioinformatics initiatives to DOE's Biofuel Centers, computing is recognized as a crucial component of cutting-edge research. MSI provides University researchers with the competitive advantage to attract these federal block grants. (See Appendix D.)
6. The trend for top-research universities is to increase support for centralized HPC resources. Among peer institutions, the University of Illinois at Urbana-Champaign is the undisputed leader in HPC, but Purdue University, the University of Michigan, and the University of Texas at Austin are dramatically increasing their support for computing resources. Over the past 15 years, the University's support of MSI has decreased from \$9.2M in 92-93 to \$6.4M in 06-07 (in nominal dollars!) with all the subsequent adverse effects in terms of prestige, visibility, and ability to be a quality research Institution. (See Appendix E.)

To help the University achieve its strategic goal to become one of the top-three public research institutions, the long-range planning committees make the following recommendations for administrative and programmatic changes to the University's high-performance computing research infrastructure and for expectations from its user community:

Structure of MSI:

- The Digital Technology Center and the Minnesota Supercomputing Institute should be restructured into a **single** unit under the leadership of a **single** director. Combining MSI and DTC will result in a cost-effective means of integrating the research, education and outreach component of the DTC with the resources and expertise of the MSI. A single center would provide the technical support and HPC resources needed university-wide, and the investigators at the DTC would drive research efforts that leverage these significant resources, building interdisciplinary bridges. This new unit should continue the use of the name "University of Minnesota Supercomputing Institute (MSI) for Digital Technology and Advanced Computation" because it carries tremendous name recognition.
- The new combined unit should have five core responsibilities: (i) digital technology research, (ii) provider of HPC hardware resources, (iii) provider of HPC software support, high-end database development and hosting, and HPC consulting services, (iv) provider of advanced visualization and satellite laboratories, and (v) HPC education and outreach. Oversight of each core area should be by a steering committee consisting of the MSI Director and four or five faculty members. The budget of the combined unit must be commensurate with the core responsibilities.

Programmatic Aspects:

- Generally, high-performance computing should not be narrowly defined. It includes demanding computational problems in all research disciplines that require enabling hardware and software resources or user support beyond those commonly found on high-end desktop computers and standard software applications. The University benefits more from MSI advancing the research of a large user group from all disciplines, than a narrow focus to traditional supercomputing research.
- MSI needs to recognize and benefit from its extremely broad user base. One-size fits all cannot work for MSI's hardware resources. While the computing needs of some users will be

met best by a massively-parallel computer with extremely fast interconnects and large memory, the computing needs of other users will be met best by a large number of fast single processors for embarrassingly parallel production runs with long wall-clock times. Both of these constitute equally valid uses of HPC resources. Thus, MSI needs to invest into various types of hardware resources, including visualization platforms.

- MSI should provide higher-level of support for a few strategic applications. The research component of MSI should target some highly visible applications that will receive computational resources and staff support at a higher level. These applications need to feature prominently in presentations to the various stake holders.
- MSI must increase its visibility and outreach. Its lecture facilities are impressive and MSI should more aggressively support international workshops focusing on any aspect of high-end computing research. MSI should start a series of monthly lectures to be presented by the University's HPC users. This will increase familiarity with other HPC areas, foster collaboration and help with team-building for block grants. MSI should improve coordination with the Institute for Mathematics and its Applications. MSI should also play a greater role in facilitating the education of HPC users. For example, federal funding agencies and industry are calling more forcefully for better verification and validation of computer algorithms and models. MSI should spearhead and coordinate such efforts that are very attractive to the funding agencies. Outreach to tribal colleges, community colleges, the Science Museum, and local high schools need to be strengthened. These efforts will help to educate the broader public about research achievements enabled by HPC and to recruit a new generation of HPC literates.
- MSI resources must be used as leverage to attract federal block grants. MSI provides the competitive advantage to attract federal centers (e.g. NIH Biomedical Computing Centers of Excellence, NSF Engineering Research Centers, and the DOE Predictive Science Academic Alliance Program). The MSI leadership must pursue these opportunities aggressively and be held accountable. The University's investment into and maintenance of an HPC infrastructure is essential to compete successfully for such block grants. However, the University must recognize that most of these block grants are unlikely to contribute directly to hardware or software resources at MSI.
- MSI provides researchers with the leverage to pursue HPC resources that are greater than those available at any academic institution. The user community must recognize that, within its current budget, MSI should not be expected to serve the hardware needs of all high-end users at all times. High-end users need to leverage expertise gained from MSI usage in order to compete for computational resources at national HPC centers (NSF Teragrid, DoD and DoE resources).
- HPC facilities and digital technology are essential infrastructure components for any research university. Thus, it would be as counterproductive to charge individual users for computing resources as for the individual download of a research article from the web version of a scientific journal or for the individual borrowing of a book from the library.

New Computing Platforms:

- MSI should explore the development of a campus computing grid to harness the huge University investment in Office and Information Technology. Stanford University has tremendous success with such a grid that now serves much of its HPC needs. A perfect test bed for such a campus computing grid would be the IT Computing Laboratories. If the experience with this grid is positive, then it may be possible to shift a larger fraction of the MSI hardware budget to high-end supercomputing to enable the University to reclaim a spot among the top quarter of the top 500 supercomputers (see www.top500.org). MSI's BladeCenter is currently ranked as 198, not even among the 50 fastest computers located at other research universities.

Other Aspects:

- The current committee structure within MSI appears to be cumbersome and limits oversight by faculty. For example, there are five separate steering and three separate executive committees for the various laboratories. These laboratories form core area (iv) of MSI and a single steering committee with broad membership should have responsibility for all satellite laboratories. This will provide improved oversight and avoid duplication of resources. The need for separate executive committees for each core area should be re-evaluated. A single executive committee can potentially handle the demands of all core areas.
- The current user account structure within MSI appears to be cumbersome and hamper efficient use of resources. MSI should establish a framework that allows a user to login to a single computer and get access onto MSI's entire computing grid. It should be set up so that certain applications when sent to the single queue are automatically run on machines that can best support this specific job. MSI should optimize this single queue according to current machine capabilities and usage. In specific cases, users may request permission priority to overrule the computing grid defaults.
- MSI user support should be re-evaluated. Could users be better served by a different distribution of support staff? Would it be helpful to list the specific expertise area of each support staff member on the web and to provide specific office hours?
- MSI should re-evaluate its space needs. Could some space be used more optimally? Faculty members do not need an office in their home department and at MSI.
- Accountability and oversight are critical for MSI to facilitate excellence in research. Awarding resources by collegiate affiliation is not in line with the University's strategic goals. Resources should be awarded competitively on the basis of scientific merit and/or the potential for success (funding, publications, personnel). Accepting MSI resources comes with responsibilities for the principal investigator. MSI should aggressively enforce proper acknowledgement of its resources in research publication and submission of these publications and other reports to the MSI Report Series. Principal investigators who fail to submit any report over a 2-year period, should not receive more than the standard resource allocation in future periods.
- MSI should establish a peer review process for support services. User support services (HPC software support, high-end database development and hosting, and HPC consulting services) are becoming increasingly important and are taking a substantial part of MSI's budget. To make prudent use of these support services, it is recommended that a peer review process be established for projects requiring significant support from MSI, e.g. more than 8 hours of consultant time.

Appendix A: Long-Range Planning Committees

At the request of Dr. Andrew Odlyzko, the MSI interim Director, two long range planning committees for the Minnesota Supercomputing Institute (MSI) were formed in October 2006. These were charged with the strategic planning of MSI support for computation in the biosciences in one case, and for the engineering, math, and physical sciences in the other. Although there exist differences in the needs and uses between the communities represented by the two committees, there are many common elements and so the decision was made early on to prepare a single report with the findings of both committees. The committees met bi-weekly during the fall semester 2006. In addition, a survey was prepared by the committees and distributed to the MSI user community. From the committee discussions and the survey results, a clear picture emerged of the vital role that MSI plays in enabling cutting-edge research at the University.

The committees consisted of the following members:

Long-Range Planning Committee for MSI Support of Computation in the Biosciences

Yiannis Kaznessis (chair), Chemical Engineering and Materials Science; Assistant Professor
Eric Celeste, Library Administrative Services; Associate Director
John A. Crow, Center for Biomedical Research Informatics; Director
Stephen C. Ekker, Genetics, Cell Biology and Development; Assoc. Professor and Associate Head
Fumiaki Katagiri, Plant Biology; Associate Professor
Arkady Khodursky, Biochemistry, Molecular Biology, and Biophysics; Assistant Professor
Vipin Kumar, Computer Science and Engineering and Scientific Computation; Professor and Head
Kevin H. Mayo, Biochemistry, Molecular Biology, and Biophysics; Professor
Claudia Neuhauser, Ecology, Evolution and Behavior; HHMI Professor and Head
Hans G. Othmer, Mathematics and Scientific Computation; Professor
J. Ilja Siepmann, Chemistry and Scientific Computation; Distinguished McKnight University
Professor and Director of Graduate Studies for Chemistry and Chemical Physics
Robert M. Sweet, Urologic Surgery; Assistant Professor
David D. Thomas, Biochemistry, Molecular Biology, and Biophysics and Scientific Computation;
Professor
Carston R. Wagner, Medicinal Chemistry; Professor and Director of Graduate Studies

Long-Range Planning Committee for MSI Support of Computation in Engineering, Math, and the Physical Sciences

J. Ilja Siepmann (chair), Chemistry and Scientific Computation; Distinguished McKnight University
Professor and Director of Graduate Studies for Chemistry and Chemical Physics
Christopher J. Cramer, Chemistry and Scientific Computation; Distinguished McKnight and
Distinguished Teaching Professor
Efi Foufoula-Georgiou, Civil Engineering; Distinguished McKnight University Professor
Timothy J. Griffin, Biochemistry, Molecular Biology and Biophysics; Assistant Professor
Thomas W. Jones, Astronomy and Scientific Computation; Professor
Yiannis Kaznessis, Chemical Engineering and Materials Science; Assistant Professor
Mitchell B. Luskin, Mathematics and Scientific Computation; Professor
Krishnan Mahesh, Aerospace Engineering and Mechanics; Associate Professor
Bob Numrich, Minnesota Supercomputing Institute: Senior Research Associate
Yousef Saad, Computer Science and Engineering and Scientific Computation; Professor
Greg Smaby, Principal, Smaby Group Inc.
Jon B. Weissman, Computer Science and Engineering; Associate Professor and Director of
Graduate Studies
Renata M. Wentzcovitch, Chemical Engineering and Materials Science; Professor

Appendix B: The Importance of Computational Sciences and High-end Computing Resources

Computing has become a *sine qua non* of cutting edge research. From engineering to neuroscience, computational research is providing scientific insight that significantly enhances and augments experimental and theoretical research endeavors. Armed with high-performance computers, it is now possible to solve previously intractable problems, quantifying phenomena and processes in a manner that far exceeds our innate capacity for theoretical, reductive reasoning.

Computing, of course, enhances our ability to quickly perform mathematical operations and compare large amounts of information. But computing does not simply enhance our intellectual and perceptual abilities or complement our mathematical talents. It augments our research endeavors by giving us instrumental access to new forms of physical intuition thus providing important stimulus for novel discovery.

Computer simulations are indeed abstracted from the material content of the system being simulated, but this is true of any other kind of theoretical model. What simulations offer is a significant reduction in the degree of idealization needed in the underlying model. Furthermore, data driven efforts are becoming an integral component of cutting-edge science, and without high-end computing resources, large data amounts storage, quick retrieval, detailed analysis, and broad dissemination are not possible.

To stress these points we can look at any scientific discipline as an example, but consider the biological and medical sciences, where a revolution is taking place. Indeed, the completion of numerous Genome Projects in the last decade marks a magnificent milestone for humankind. This ongoing progress is, however, merely the harbinger of a revolution in biological sciences. The deciphering of principles that guide nature to express itself from DNA to protein sequence, to structure and function and then on to complex interactions and phenotypes will not only provide us with deep understanding of the secrets of life, but will also allow us to control such biological phenomena as cell differentiation and cell death. Besides the human genome, the genomes of more than 165,000 different organisms have been already completely or partially sequenced with more 100,000,000,000 bases now known (see Press Release from the National Center for Biotechnology Information http://www.nlm.nih.gov/news/press_releases/dna_rna_100_gig.html), and this volume of genomic data will only continue to grow exponentially over the next years.

The challenge of generating knowledge from such unprecedented volumes of data is compounded by the complexity encompassed in these datasets. It is evident that the paradigm of biological sciences as descriptive disciplines cannot rapidly assist in rationally engineering diagnostic and therapeutic technologies. Thus, the task facing the scientific and engineering communities is to reduce this enormous volume and complexity of biological information into concise quantitative formulations with predictive ability, which in turn will be useful for medical and biotechnological applications. This is indeed an immense task and progress relies heavily on HPC.

In practice, it is now possible to develop models of complexity beyond any human comprehension capability. Computer models of protein folding, biomolecular interaction dynamics, tissue mechanics, organismal and ecological behavior can encompass a much larger number of variable degrees of freedom than any human brain can. Through the solution of these models in time and space, physics, phenotypes and behaviors emerge that cannot be captured by any thought experiment or paper and pencil techniques. Furthermore, insight is gained that cannot be distilled by any set of experiments.

It is easy to see how similar arguments can be made for all scientific and engineering disciplines, e.g. in materials design, process engineering, astrophysics, aerospace and automotive sciences.

In conclusion, high-end computing is not simply a convenient resource to have. ***It is the only means for conquering new scientific discoveries and exploring research avenues,*** accessible with no other way.

Appendix C: Summary of MSI User Survey

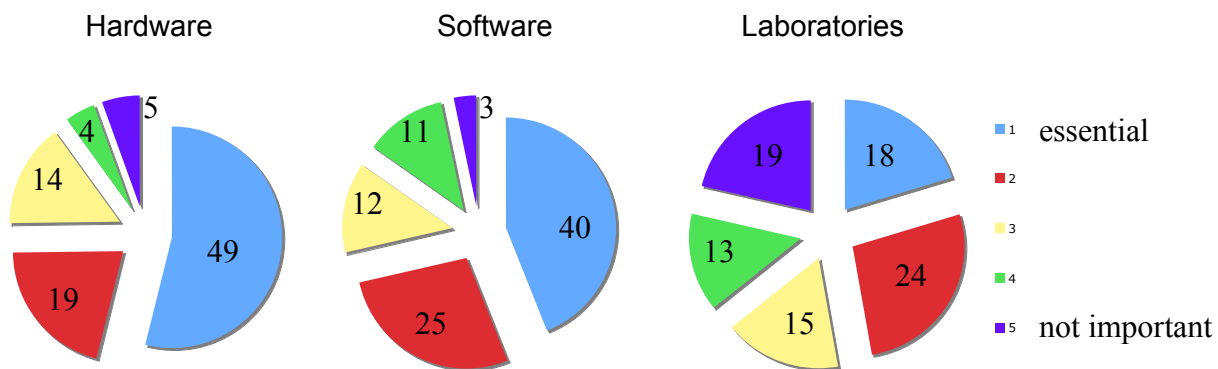
To assess the role of MSI, an internal survey was conducted. More than 300 University of Minnesota principal investigators are MSI resource users. 91 PIs responded to the survey that was sent out in December of 2006, a remarkable response given that only two weeks were available for PIs to respond. Complete survey questions and the responses are available from MSI.

The user community has overwhelmingly responded in a positive way to the survey submitted by the committees. The vast majority of users states that MSI resources are essential for conducting research. Hardware, laboratories and user support are the major elements of MSI resources that enable University faculty to advance their work.

The first few questions are related to the heart of the matter:

1. How important are MSI's hardware resources (processor/memory/storage) to your research?
2. How important are MSI's software support, database development and hosting, and consulting services to your research?
3. How important are MSI's laboratories (CGL, SDVL, SDML, BSCL, VWL) for your research?

The responses are as follows: [1 = essential; 5 = not important; # of responses on pie-chart]



In the detailed comments elicited from the user community we see an overwhelming positive response to the following question "Describe the importance of MSI resources for your research endeavors (e.g. solving scientific problems, attracting funding, attracting graduate students and postdocs)."

Representative responses are:

- MSI is of the utmost important for solving most of our scientific problems (about 80% of our research publications) and it greatly helps in attracting graduate students and postdocs. Without the success in research supported by MSI, we would not be able to attract funding.
- The resources provided have been essential for the successful completion of several research projects, which in turn will serve as the basis for future proposals. Results obtained to date have already been used to generate and maintain the interest of several industrial sponsors. Relatively easy access to these resources has also helped retain and attract excellent graduate students.
- MSI is entirely important to the advance of the proteomic program at the University
- Critical. We use the phylogenetic software packages extensively and mention their availability in every grant proposal. I have mentioned the MSI to potential grad students, postdocs, and importantly, potential new faculty. Postdocs and faculty recruits always ask about computing

resources.

- MSI is absolutely critical to our research efforts, to my ability to recruit effectively, and when I travel around world collaborators and competitors all know the name, admire the resource, and wish that they had something similar. I see the Bulletin sitting on desks in the most unexpected places. My ability to maintain grant support derives from our productivity, which is vastly greater with MSI than it would be otherwise. Given the typical funding (in \$) of grant agencies like NSF for individual workstations, we could never purchase in-lab resources that would be adequate to our needs. And, national centers are structured in a way that makes them vastly less convenient to use (long waits, poor software support, etc.) than MSI.

The responses are clearly positive, demonstrating that MSI's role is not simply adding convenience to the user community. Instead it is a critical resource, not available elsewhere. It also became evident during the committees discussions that such resources are not offered as a remote resources and cannot be beneficial as fragmented resources across campuses and Departments. The MSI resources in terms of high-end CPU cycles, data storage and analysis, software, technical and user support constitute the necessary critical mass for promoting excellence in research without duplication of efforts and costly fragmentation of resources.

Perhaps a critical question is related to the return on investment. Although efforts have been made in the past to associate external funding to MSI resource availability, the total dollar amount of federal grants awarded to all MSI users may lead to misleading conclusions and be open to criticism. Instead, in the survey the MSI users were asked the following question: "Please specify the \$ amount of your external funding that is directly supporting research efforts utilizing MSI resources." We clarified this statement further by saying that if for example the user PI has an external grant supporting two graduate students and the students have a 40% research effort solely utilizing MSI resources, only 40% of the grant should be included in the response.

Based on the responses to this survey, we find that in 2006, more than \$13 million in federal funds awarded to the 91 respondents were associated *directly* with research efforts expended at and aided by MSI resources. One may extrapolate with caution from these 91 responses to the entire user base of MSI, and it appears likely that about \$40 million of the \$90 million of research grants that include high-performance computing in the work plan, is devoted to research activities that do involve MSI directly. This response unambiguously demonstrates that ***MSI provides a direct benefit to the University in terms of funding, recognition, prestige, research productivity, and ability to attract and retain good people.***

Another clear conclusion drawn from our discussions and the survey results is that local, centralized computing resources are a cost effective way for nucleating strengths and expertise that enable science. Centralized resources provide the critical mass that is crucial for a wide spectrum of research efforts, from quantum chemistry to clinical informatics. The absence of funding mechanisms for small scale computing resources would put in jeopardy the work of scientists needing computing cycles, scientists needing data storage resources and scientists needing expensive software and software support. Furthermore, distributed resources would result in increased costs for hardware, software, maintenance and security support. Start-up packages for researchers using HPC would also increase substantially, as well as requirements for space and support personnel. In addition, while the throughput of data networking has increased at a phenomenal rate, the quantity of data being generated by researchers is growing at an even greater rate. Processing data remotely presents a significant cost in the time it takes to transfer data to and from the remote systems. This affects researchers who are dealing with huge data sets and also affects researchers who are generating data that needs to be visualized interactively.

The concerns that were raised by the community were overwhelmingly related to the adequacy of resources. ***The user community voiced a strong concern regarding the troubling trends of diminishing investments to MSI and strongly supports a strong MSI as a centralized resource of high-end computing infrastructure and support.***

Appendix D: The Role of High-end Computing Supported Research as seen by the Federal Funding Agencies

The major funding agencies (NIH, DOE, NSF, DOD) are formulating strategic visions for scientific, clinical and engineering discoveries that are inherently based on computational sciences and large, centralized computing resources.

In January 2003 the National Science Foundation (NSF) issued the report from the Blue-Ribbon Advisory Panel on Cyberinfrastructure, which was chaired by Professor Daniel E. Atkins of the University of Michigan (<http://www.nsf.gov/od/oci/reports/toc.jsp>). This is indeed a remarkable document that should be consulted by everyone involved in academic decision making. The report, entitled Revolutionizing Science and Engineering Through Cyberinfrastructure, noted that:

“The Panel’s overarching finding is that a new age has dawned in scientific and engineering research, pushed by continuing progress in computing, information, and communication technology; and pulled by the expanding complexity, scope, and scale of today’s research challenges. The capacity of this technology has crossed thresholds that now make possible a comprehensive “cyberinfrastructure” on which to build new types of scientific and engineering knowledge environments and organizations and to pursue research in new ways and with increased efficacy. The cost of not doing this is high, both in opportunities lost and through increasing fragmentation and balkanization of the research communities”.

“The Panel’s overarching recommendation is that the National Science Foundation should establish and lead a large-scale, interagency, and internationally coordinated Advanced Cyberinfrastructure Program (ACP) to create, deploy, and apply cyberinfrastructure in ways that radically empower all scientific and engineering research and allied education.”

Following publication of the report from the Blue-Ribbon Advisory Panel, reports from all of NSF’s directorates have stressed the need for a national cyberinfrastructure for science and engineering with tools, services, and resources far beyond those currently available. The NSF’s 2005 budget request included \$137.9 million for the widely shared cyberinfrastructure program and continued investments are supported in the 2006 and 2007 budgets. Academic institutions have been positioning themselves to benefit from these available resources. For example, the University of Texas competed successfully for a large NSF Cyberinfrastructure grant (\$59 million) only after significant internal investments to their high-performance computing center.

The National Institutes of Health place an increasingly important role on computational biosciences. In the NIH Roadmap for Medical Research, Bioinformatics and Computational Biology play a prominent role. NIH’s Director Dr. Elias Zerhouni notes the following (from <http://nihroadmap.nih.gov/>):

“Today’s biomedical researcher routinely generates an amount of data that would fill multiple compact discs, each containing billions of bytes of data. (A byte is approximately the amount of information contained in an individual letter of type on this page.) There is no way to manage these data by hand. What researchers need are computer programs and other tools to evaluate, combine, and visualize these data. [...] By embarking on the Bioinformatics and Computational Biology initiatives, the NIH Roadmap is paving a future “information superhighway” dedicated to advancing medical research. A central focus of the effort will be a set of National Centers for Biomedical Computing.

As the Centers begin to generate the software and data management tools to serve as fundamental building blocks for 21st century medical research, individual scientists will be funded to work together with the centers. “Big science” and “small science” will work hand-in-hand to advance all of biomedical research. Through these efforts, researchers will be able to share data gathered from large experiments. The best minds will be able to work together effectively to tackle unsolved mysteries, such as the role of heredity in individuals’ different responses to medicines

and the complex interplay of genetic and environmental factors in common diseases such as Alzheimer's disease, heart disease, cancer, and diabetes.

The Bioinformatics and Computational Biology initiatives will create a national software engineering system. Through a computer-based grid, biologists, chemists, physicists, computer scientists, and physicians anywhere in the country will be able to share and analyze data using a common set of software tools. Developers of the project envision that the system will resemble that of the integrated software packages for office tools installed on most home computers today, in which information can be traded seamlessly between software such as spreadsheets, word-processing and e-mail programs."

In 2004, four centers, part of the National Institutes of Health Roadmap for Medical Research, were funded. They are supported by 5-year grants projected to total more than \$79.7 million (\$15.7 million the first year). Another three centers were funded in 2006 for a total of seven.

The Department of Defense and the Department of Energy have added high-computing based research a major components of their strategic discovery vision. The University of Minnesota benefits from DoD commitment to HPC through the Army High Performance Computing Research Center. The University also benefited from MSI in the effort to submit a proposal for the Department of Energy Genomics to Life (GTL) Bioenergy Research Center. Computational systems biology research and supporting resources are a key element of the DOE Biofuel Center white paper and call for proposals. From the call for proposals the overview reads:

"The mission of the Bioenergy Research Centers will lie at the frontier between basic and applied science, and will maintain a focus on bioenergy applications. The purpose of the Department in funding the Centers will be to achieve real steps toward real solutions to the challenge of producing renewable, carbon-neutral energy. At the same time, it is desired that the Centers be grounded in basic research, pursuing alternative avenues and a range of high-risk, high-return approaches to finding solutions."

The description of the GTL includes the following (<http://genomicsgtl.energy.gov/compbio/index.shtml>):

"An Essential Foundation: Computation is essential to the GTL program goal of achieving a predictive understanding of microbial cell and community systems. Computing and information technologies allow us to surmount the barrier of complexity that separates genome sequence from biological function. The integrated GTL computational environment will link data of unprecedented scale, complexity, and dimensionality with theory, modeling, simulation, and experimentation to derive principles and develop and test biosystems theory. GTL computation will employ data-intensive bioinformatics, compute-intensive molecular modeling, and complexity-dominated cellular systems modeling. Models and simulations represent an ultimate level of integrated understanding. A key goal for cell modeling is to predict cell phenotype from the cell's genotype and extracellular environmental information. Such predictions, resulting from comparative genomics studies, will include cell ultrastructure, morphology, motility, metabolism, life cycle, and behavior under a wide range of environmental conditions. These models not only will be descriptive and phenomenological but also will be predictive at multiple levels of detail."

With MSI expertise and resources, the University developed a credible plan for 'Big Science' in biofuels research. Any plan would be substantially weakened without a centralized high-end computing facility at the University of Minnesota. Indeed, if Minnesota's biofuels research proposal is funded, then MSI will receive funds to hire several technical employees and to manage their work.

Appendix E: High-end Computing Initiatives at Peer Institutions

Looking at other Institutions, there is a clear trend that being stewards of the scientific and engineering disciplines is associated with strong support of computational sciences and continued investments in considerable HPC resources. Concentrating on Big 10 schools, there are clear trends of increasing investments for HPC resources and computational sciences.

The **University of Illinois at Urbana-Champaign** is the undisputed leader in HPC, with the renowned National Center for Supercomputing Applications (NCSA) holding the helm of supercomputing centers nationwide since 1986. Continued support from the state of Illinois and UIUC has resulted in secured funding from the NSF for over two decades. The truly astonishing resources at NCSA are described in the most recent annual report (http://www.ncsa.uiuc.edu/AboutUs/NCSA_Annual_Report_05.pdf), where Professor Thomas H. Dunning Jr., the Director of NCSA pledges continued provision of the leading-edge computing, data, and visualization resources to power cyberenvironments for science and engineering.

In the last few years, **Indiana University and Purdue University** made significant investments in high-end computing resources that paid-off with a \$30 million award from NSF for IU and PU to become a TeraGrid site. "TeraGrid is an open scientific discovery infrastructure combining leadership class resources at nine partner sites to create an integrated, persistent computational resource" (from <http://www.teragrid.org/about/>). "Using high-performance network connections, the TeraGrid integrates high-performance computers, data resources and tools, and high-end experimental facilities around the country. These integrated resources include more than 102 teraflops of computing capability and more than 15 petabytes of online and archival data storage with rapid access and retrieval over high-performance networks. Through the TeraGrid, researchers can access over 100 discipline-specific databases. With this combination of resources, the TeraGrid is the world's largest, most comprehensive distributed cyberinfrastructure for open scientific research." Exciting discovery efforts are enabled from these HPC resources (from <http://discoverypark.purdue.edu/wps/portal/Cyber>): "The Cyber Center is creating a human infrastructure for collaboration and research for projects engaging cyberinfrastructure at Purdue. The goal of the Cyber Center is to create a center of national preeminence in computational methods for discovery and learning. Cyberinfrastructure is infrastructure based upon distributed computer, information, and communication technology. The Cyber Center's vision is to create an interdisciplinary facility of national prominence that will: engage in basic cyberinfrastructure research, develop new cyberinfrastructure tools and techniques, and deploy results to real communities meeting real user needs at Purdue and beyond."

At the **University of Michigan**, the Center for Advanced Computing (CAC) delivers high performance computing, grid infrastructure, very large data storage, and advanced visualization services through the College of Engineering and throughout the University of Michigan. With over 1000 processors, two high-speed switches for fast networking and terabytes of storage, the CAC can provide resources for most any computation problem size (from <http://cac.engin.umich.edu/>). Furthermore, the Michigan Center for Biological Information provides considerable HPC resources to bioscience, clinical and medicinal researchers.

Finally, **Iowa State** built recently its first high-end computer with both internal funding and funding from NSF. In a Press Release, Gregory Geoffroy, the President of Iowa State University said "Iowa State University's first supercomputer will be a tremendous resource for our researchers involved in assembling the corn genome, studying protein networks and other very important areas of computational biology. It will help us strengthen our programs in the biological and information sciences and support our faculty as they create, share and apply knowledge to benefit Iowa and the world." (from <http://www.iastate.edu/~nscentral/news/06/jan/supercomputer.shtml>)