Establishing a Compact Muon Solenoid Grid-based Tier-2 National Computing Center and NSF Shared Cyberinfrastructure Tier-2C Program at The University of Iowa

A Proposal to Establish a
Grid-based Tier-2 National Computing Center for the Large Hadron Collider Compact Muon Solenoid Experiment and National Science Foundation Shared Cyberinfrastructure Tier-2C Program at The University of Iowa

List of all Personnel Associated with the Proposal

Principle Investigators	Senior Personnel	
Gregory R. Carmichael[3, 7], Co-PI	Kurt M. Anstreicher[9, 12]	Edwin Norbeck[15]
Yasar Onel[15], PI	Marc P. Armstrong[3, 11, 20]	Suely Oliveira[9]
Ge Wang[4, 8, 13, 16], Co-PI	David A. Bennett[3, 11]	Virendra C. Patel[2, 14]
Shaowen Wang[11, 19], Co-PI, Technical Director	Thomas L. Casavant[1, 5, 10]	Jerald L. Schnoor[3, 8]
	Mary Kathryn Cowles[6, 17]	Alberto M. Segre[9, 21]
	Steven R. Fleagle[19, 22]	Luke Tierney[17]
	Boyd M. Knosp[19]	Mark A. Wilson[18]
	Jun Ni[9, 19]	Jun Yan[17]

Affiliations

[1]Center for Bioinformatics and Computational Biology (CBCB)
[2]Center for Computer-Aided Design (CCAD)
[3]Center for Global and Regional Environmental Research (CGRER)
[4]Center for X-Ray & Optical Tomography
[5]Department of Biomedical Engineering
[6]Department of Biostatistics
[7]Department of Chemical and Biochemical Engineering
[8]Department of Civil and Environmental Engineering
[9]Department of Computer Science
[10]Department of Electrical and Computer Engineering
[11]Department of Geography
[12]Department of Management Sciences
[13]Department of Mathematics
[14]Department of Mechanical and Industrial Engineering
[15]Department of Physics and Astronomy
[16]Department of Radiology
[17]Department of Statistics and Actuarial Science
[18]Iowa Institute of Hydraulic Research (IIHR)
[19]Information Technology Services
[20]Program in Applied Mathematical and Computational Sciences
[21]Program in Public Health Genetics
[22]Vice President for Research Office

The University of Iowa
Iowa City, IA 52242

October 29, 2004

Project Summary: Establishing a Compact Muon Solenoid Grid-based Tier-2 National Computing Center and NSF Shared Cyberinfrastructure Tier-2C Program at The University of Iowa

Scientific and engineering endeavors have become dependent on analyses that require shared cyberinfrastructure to support collaborative problem solving. The Grid Research and educatiOn group @ IoWa (GROW) is responding to the challenge of creating this necessary cyberinfrastructure by proposing to establishing a Compact Muon Solenoid Grid-based Tier-2 national computing center and NSF shared cyberinfrastructure Tier-2C program at The University of Iowa. Based on our proof-of-concept CMS Tier-2 experience, our project will advance user- and application-level the Open Science Grid (OSG) capabilities by creating a production interface between OSG and our campus Grid infrastructure – HawkGrid.

Our proof-of-concept CMS Tier-2 experience has demonstrated that GROW successfully acted as a physics Grid community member through collaboration with other CMS physics Grid infrastructure teams and delivery of rapid and reliable Tier-2 level services. Through this experience, GROW has also become an effective university community effort to Grid-enable a number of domain scientific and engineering applications other than from high energy and nuclear physics and to carry out education and outreach activities.

We aim to serve the entire CMS collaboration as a production Tier-2 center with full computing and storage capacities supported by a strong university commitment and the forward-thinking vision of GROW-affiliated researchers. Our Tier-2 services will center on CMS data storage, simulation and analysis, research and deployment of interoperable Grids, and research and development of user and application level OSG capabilities. The services for providing user- and application-level OSG capabilities will benefit a variety of cyberinfrastructure projects in our Tier-2C program that will, in turn, enhance and strengthen these Tier-2 services through the GROW framework. Our strong university support includes a strategic plan to upgrade our external network to the level (2.5-10GigB/s) required to operate a Tier-2 center before 2007. On-going network testing between Fermi Lab and our proof-of-concept Tier-2 systems has demonstrated that an existing 1Gb/s link to our systems consistently achieves the best performance in two of six network tests among all universities and national labs that participate in this evaluation experiment. The GROW vision resonates with the OSG strategic plan with respect to developing a persistent Grid infrastructure that is shared across scientific and engineering communities by federating and empowering university community resources.

Our Tier-2C project will be guided by this vision, organized by the GROW knowledge sharing and management framework, and supported by the HawkGrid campus grid infrastructure. HawkGrid technologies will continue to be OSG-compatible to support the federation of campus resources that are opportunistically available to CMS. The experiences of developing community-specific knowledge environments for research and education will be shared effectively in the Tier-2C program through a Grid application team, while interdisciplinary research will be fostered through a science committee. Moreover, these experiences will be used to continually improve our Tier-2 services, and will be shared with CMS collaborators through regional and national workshops such as the one held at The University of Iowa this past summer.

Intellectual Merit: Our Tier-2 center and Tier-2C program are focused on creating new OSG capabilities, integrate, and deploy stable OSG capabilities to empower individuals, groups, universities, and the entire scientific and engineering community. GROW implements this focus by delivering reliable Grid infrastructure services that bridge scientific and engineering domain communities using Grid-based cyberinfrastructure (on our campus and beyond) through our participation in and collaboration with OSG. The effects of our Tier-2/Tier-2C will be profound. The technical and scientific challenges are significant in operating and advancing Grid-based cyberinfrastructure as well as building cyberinfrastructure-based knowledge environments for research and education. Through strong and organized community effort, we have assembled a team that has the needed breadth and depth to tackle significant operation and research challenges.

Broader Impact: The effects of the proposed Tier-2 center and Tier-2C program will be far reaching. First, both will advance cyberinfrastructure development and enable CMS scientific discovery that promises to break new ground in understanding the nature of the universe and its fundamental building blocks. Second, they will facilitate the development of our shared cyberinfrastructure and empower our university computational community to engage in collaborative problem solving that is computationally intensive. 9 disciplines from 22 academic units on campus participate in and contribute to the Tier-2C program. Third, our Tier-2 and Tier-2C effort will promote and facilitate cyberinfrastructure research and education on The University of Iowa campus, in the Midwest region, and beyond. For example, a new university funding initiative (see the support letter from The University of Iowa Office of Vice President for Research) will be created to support new cyberinfrastructure research and education projects that would make use of our proposed Tier-2 center and Tier-2C program.

Establishing a Compact Muon Solenoid Grid-based Tier-2 National Computing Center and NSF Shared Cyberinfrastructure Tier-2C Program at The University of Iowa

The University of Iowa (UI) Grid Research and educatiOn group @ IoWa (GROW) hereby submits a proposal to become a USCMS Tier-2 and National Science Foundation shared cyberinfrastructure Tier-2C center. This center will leverage existing UI activities and resources and external funds to create a powerful entity in support of the goals of the CMS experiment and advancement of Grid-based shared cyberinfrastructure. The proposed effort, which will be carried out by GROW (see V. Management Plan), represents a broad spectrum of expertise in CMS data analysis and Grid computing as well as the computer and computational, sciences. GROW is based on the UI information technology services (ITS) with its mission to enable domain science and engineering through OSG cyberinfrastructure by strengthening and broadening collaborations between the Grid computing community and the domain sciences and engineering communities (Figure 1). This proposed center will take advantage of current UI facilities and resources that include a proof-of-concept CMS Tier-2 center and extensible campus grid – HawkGrid. This proposal is comprised of seven main parts: Tier-2, Tier-2C, Education and Outreach, Broader Impact, Management Plan, Budget, and Budget Justification.

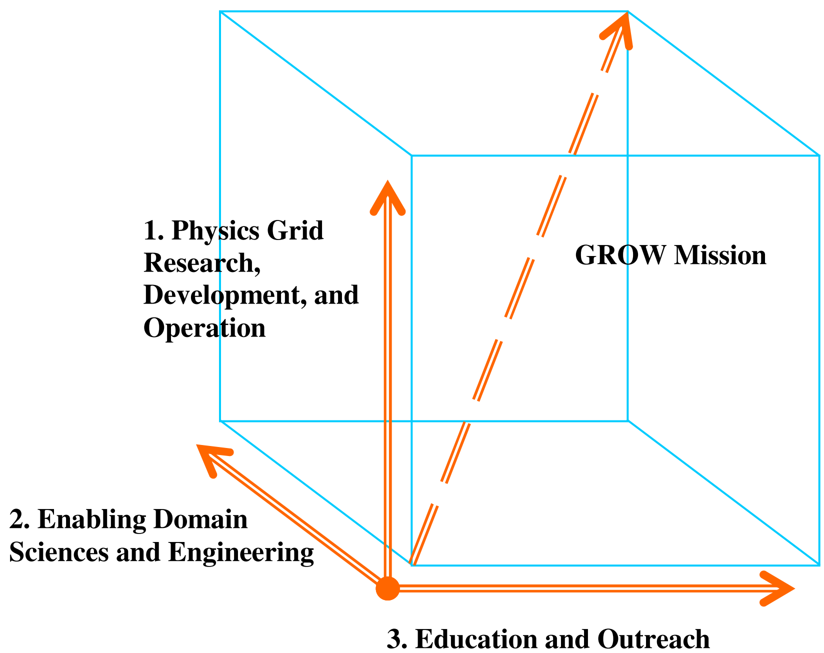


Figure 1. GROW mission

I. The University of Iowa Tier-2: An Open Science Grid-based National Computing Center for Large Hadron Collider Compact Muon Solenoid Experiment

This part of the proposal addresses the development of future UI CMS Tier-2 center based on the current proof-of-concept Tier-2 experience. Section 1 describes the proof-of-concept CMS Tier-2 center established at UI. Section 2 summarizes high energy and nuclear physics research and education effort as well as current CMS activities at UI. Section 3 delineates the specific personnel and computing resources that have contributed to the proof-of-concept Tier-2, and will continue to contribute to the development of the proposed Tier-2. Section 4 provides responses to the criteria described in the Request for Proposals. This section indicates the extent to which the proposed Tier-2 center will leverage existing UI activities and resources that will play an important role in the execution of this project. Section 5 briefly describes the center operation model.

1. UI Proof-of-Concept CMS Tier-2 Center

Several years ago, UI submitted a letter of intent to the USCMS Software and Computing Project to establish a Tier-2 center at UI. Since then, significant support has been received from the UI Vice President for Research Office, ITS, and the USCMS Software and Computing Project to create a proof-of-concept Tier-2 center. This center has enabled UI researchers to participate in a wide variety of CMS software and computing projects. Such projects include developing and deploying USCMS DGT (Development Grid Testbed) and the iVDGL (international Virtual Data Grid Laboratory), deploying CERN LCG (LHC Computing Grid) software, developing the OSG consortium, and submitting the proposals for the NSF ITR projects: DAWN (Dynamic Analysis Workspace with kNowledge) and GECSR (A Global Grid-Enabled Collaboratory for Scientific Research).

This section describes the overall goals of the UI proof-of-concept CMS Tier-2 center. These goals have been successfully accomplished through the execution of several important tasks. These tasks will be further developed toward production Tier-2 services. Executing the tasks has helped establish a long-term Grid research and development roadmap that will be useful to support the operation of the proposed Tier-2.

1.1. Goals

The overarching goals of the UI proof-of-concept CMS Tier-2 center are to:

- Research, develop, and deploy Grid computing technologies used for CMS data handling and analysis;
- Collaborate with the Tier-0, Tier-1, prototype Tier-2 centers, and other CMS groups to validate the LHC tiered distributed computing model;
- Evaluate hardware solutions used by the Tier-1 and prototype Tier-2 centers and share the evaluation results with these centers;
- Conduct proof-of-concept Grid-based CMS physics data analysis from local, regional, and national users;
- Introduce OSG technologies to other domain sciences and engineering for computationally intensive problem solving; and
- Conduct education and outreach.

These goals have been successfully accomplished as reflected by the following section that describes the major tasks carried out by the center. These tasks will continue to be part of the proposed Tier-2 center.

1.2. Major Tasks of the Center

This section describes the particulars of the major tasks of the UI CMS proof-of-concept center.

1.2.1. Configuration and Operation of a USCMS DGT system

The execution of this task is based on an Athlon Linux cluster (see Section 4.a) that has been used to experiment with Virtual Data Toolkit (VDT) and USCMS Distributed Processing Environment (DPE) software. This cluster has played a Development Grid Testbed (DGT) site role in USCMS Grid testbeds since the spring of 2003. It was also used to participate in a Supercomputing 2003 Conference demo led by the University of Florida prototype Tier-2 team (Figure 2).

1.2.2. LCG2 Operation

The configuration and operation of the LCG2 system is based on a Xeon cluster (see Section 4.a). Our effort is specifically focused on working with Fermi Lab to investigate the interoperability between LCG2 and VDT-based DPE. The LCG2 Resource Broker (RB) and the Berkeley Database Information Index (BDII) have been extensively studied to evaluate the ways that they are used to find the best configuration of Grid resource for jobs. We conducted a successful experiment using a LCG packaging mechanism with CMKIN_3_1_0 which was configured as an LCG2 managed program. Also, OSCAR 2.4.5 was studied during several experiments in which DARballs was used at our site. In the aggregate, in the
ollowing aspects:

- Resource Broker;
- DAR;
- RLS;
- CMKIN;
- OSCAR; and
- Sandboxing,

extensive skills have been learned to support interoperability between LCG2 and VDT-based DPE. For example, the data files for CMKIN program are known before the program is run. Consequently, they can be specified in the job description language. However, an OSCAR output file is not known until runtime. Therefore, we designed a "mechanism" that allows the OSCAR file to be registered after the program has terminated.

1.2.3. Participation in Developing OSG and LHC collaboratory

The center has actively participated in designing future OSG-based dynamic workspaces and a LHC collaboratory. We were one of the USCMS iVDGL institutions, and we also contributed to the development of DAWN (Dynamic Analysis Workspace with kNowledge) and GECSR (Grid-Enabled Collaboratory for Scientific Research) proposals to the NSF Information Technology Research program.

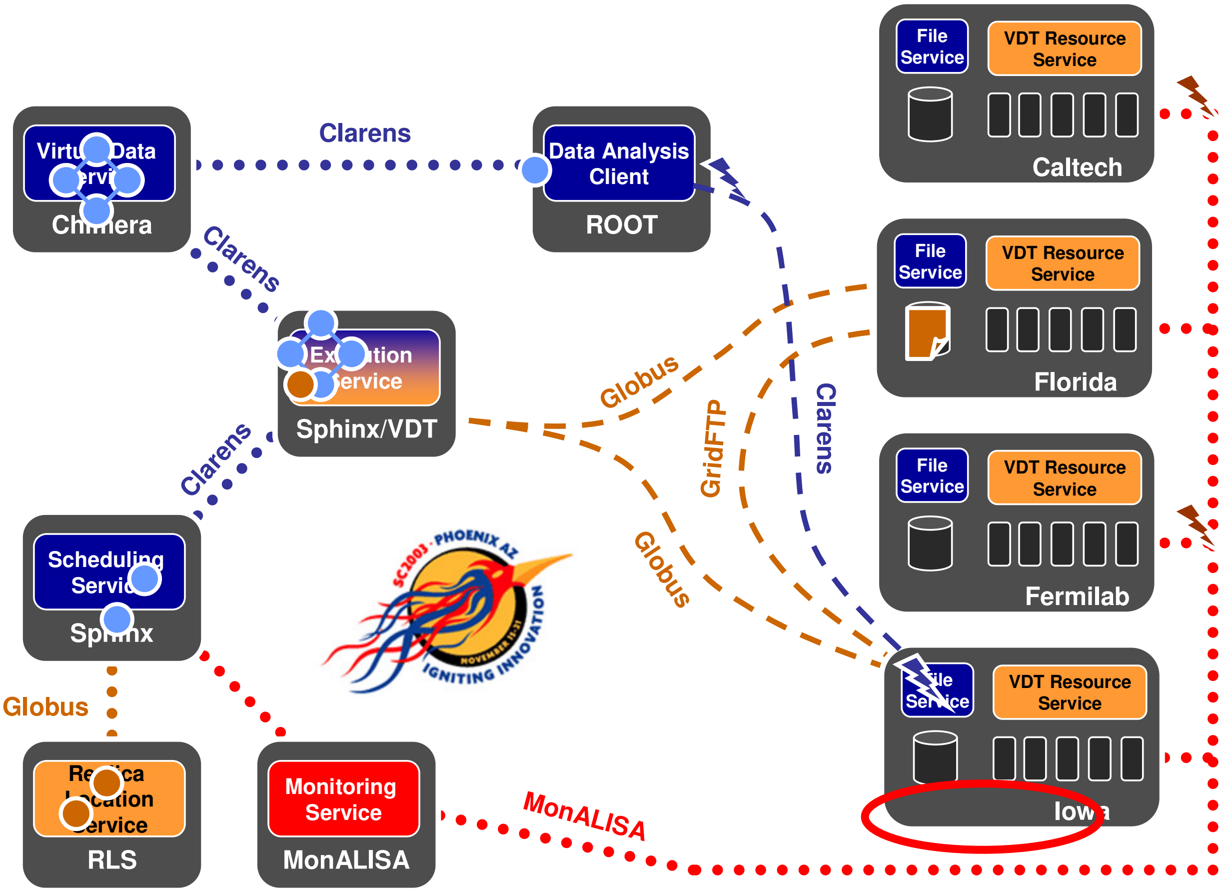


Figure 2. A Supercomputing 2003 Conference demo – Grid-based CMS data analysis, from Cavanaugh (2003)

1.2.4. Tools Providing an Interface between CMS Data Handling and Analysis

We have been working with Fermi Lab to develop tools that provide an interface between CMS data handling and analysis through the CMS Detector Database (DetDB) project. We collaborate with groups managed by Shuichi Kunori and Lee Lueking, respectively, to provide database expertise and participate in DetDB tools and applications development. A detector database stores information about detector equipment, construction, configurations, and conditions, as well as CMS experiment data. The DetDB project is building database applications for CMS HCAL, EMU, and PIXEL test beam experiments to support test data collection, detector and experiment management, data retrieval and browsing, and database access for CMS analysis tools. When CMS experiments commence in 2007, DetDB will provide full scale database applications at production level for both Fermi Lab and CERN. Both Fermilab and CERN use the Oracle database, on which most of our database applications are built. Our short-term goal is to develop a database access interface to Oracle databases for user-level analysis tools, such as ROOT. We work closely with the Fermi Lab CMS group to provide immediate Oracle database access for ROOT applications. For generic Oracle database access, we work with the Database Application group on implementing RDBC, the new database access specification of ROOT. After that, an object-oriented data model will be built to support object views in ROOT over relational databases, which will facilitate ROOT database applications development processes. In the long run, DetDB will support database access in the CMS Grid analysis environment. Such needs include Grid-level data replication, transfer, and analysis. With our knowledge of Grid computing, we wish to provide technical and research support in future large-scale and distributed database application developments for CMS data analysis as a Tier-2 center.

1.2.5. Interoperability between OSG and other Grid Environments

In-depth knowledge has been gained in understanding interoperability between OSG and other Grid environments through simultaneously working on TeraGrid applications, OSG technologies, and the LCG testbed. The operation of a LCG2 system is described in section 2.2.2. Also, our center is actively involved in developing two TeraGrid applications. These two applications are able to use TeraGrid resources and our physics Grid resources in a seamless way from the perspective of an application-level resource scheduler.

1.2.6. Dissemination of Knowledge about the USCMS Tier-2 Program and OSG Technologies
The center has provided leadership in dissemination of information and knowledge about the USCMS Tier-2 program and OSG technologies within a regional context. We are closely interacting with several Midwest Universities including Iowa State University, Kansas State University, University of Kansas, University of Nebraska, and several colleges in Iowa. We have successfully held a regional workshop[1] at The University of Iowa to:

1.) relate to participants our proof-of-concept USCMS Tier-2 center experience;
2.) share ideas on the development of a production Tier-2 center at The University of Iowa; and
3.) discuss the feasibility of creating a Midwest sub-consortium that would increase awareness and deployment of OSG throughout the region.

This workshop compliments the Caltech Tier-2 workshop in terms of clarifying what physics analysis users can contribute to advance user- and application-level OSG capabilities.

1.2.7. Enabling other Domain Sciences and Engineering to Apply OSG technologies
The center has developed a wide spectrum of collaborations with other domain scientific and engineering communities. These collaborations have significantly helped apply OSG technologies to solve problems from various disciplines (see Reference highlighting those authors who are GROW members). This task will be further elucidated in the Tier-2C part of this proposal.

1.2.8. Education and Outreach Program
A flexible and versatile education and outreach (E & O) program has been established. The E & O program of the center leverages necessary resources from three existing E & O programs: Iowa QuarkNet, Science Education Center of College of Education, and the Studio of Academic Technologies for Education. For example, Iowa QuarkNet resources have been leveraged to teach high school students about Physics Grid technologies through hands-on experience. The resources from these existing E & O programs have been assembled as needed to accomplish specific E & O goals.

The tasks described above have been carried out with significant UI support. These tasks are designed with a long-term goal in terms of providing stable Tier-2 level services. In parallel to the execution of these tasks, the center has also developed a strategic roadmap for conducting Grid computing research.

1.3. Grid Research and Development Roadmap
A long-term Grid research and development roadmap has been created not only to support CMS data handling and analysis, but also to enable use by other domain sciences and engineering. Existing Grid capabilities often focus on resource-level functions such as those for resource management. There is a significant gap between resource-level functions and what users and applications require to utilize the Grid for data handling and analysis. For example, MonaLisa[2] has created powerful capabilities for monitoring resources. However it is still not straightforward to directly use these capabilities in a resource broker implementation, even for a single application. Condor-G and Condor-C[3] attempt to provide a user-level interface to Grid computing environments. But these tools are still immature when used to provide advanced user-level Grid capabilities such as flexibly tuning the performance of resource brokering that is adaptive to the characteristics of application and resource dynamics.

Our roadmap addresses this gap by developing user- and application-level Grid capabilities with a triangle approach that investigates user-level resource brokering, a hybrid method for resource monitoring and discovery, and Grid service technologies in an aggregated way. Each aspect of this approach is closely related to the other two aspects.

1.3.1. Resource Monitoring and Discovery
We have prototyped a user-level resource monitoring toolkit that is based on multi-agent systems and geographic information systems (Wang *et al.*, 2003). This toolkit provides a rich set of APIs (Application Programming Interfaces) for embedding resource monitoring functions into end-user Grid-based applications. More importantly, this toolkit has provided a component-based environment for studying both centralized and decentralized approaches to resource monitoring and discovery.

We are currently developing and evaluating a hybrid method for Grid resource discovery that gains strength from both centralized and decentralized approaches. An unstructured peer-to-peer network is created to eliminate the problem of unique IDs as well as to address scalability. Then we use gossip as a mechanism to identify computational resources with similar characteristics and to group them. The number of messages gossiped is kept to a minimum to conserve bandwidth (Kermarrec

[1] http://www.physics.uiowa.edu/tier2/
[2] http://monalisa.cacr.caltech.edu/
[3] http://www.cs.wisc.edu/condor/

et al., 2003). Each group is represented by a leader node. All the members in the group will publish their resource availability information to this leader node. Thus, a leader node can be thought of as a local server that provides matchmaking service at a group level. Preliminary results have demonstrated that this hybrid approach is able to handle a large number of Grid resources, and is scalable and efficient with respect to lookup performance (Padmanabhan *et al.*, 2004).

1.3.2. User-level Resource Brokering

Virtual organization-level resource brokering often considers resource sharing policies, resource availability, and resource use optimality. However, user-level resource brokering is always focused on the quality of service from application needs as well as optimal selection and reservation of available resources. We investigate user-level resource brokering strategies that take advantage of the hybrid method of resource monitoring and discovery. Currently, the investigation uses our LCG2 system to conduct experiments for developing and evaluating optimization-based resource matching methods. Future research will continue to benchmark these methods and develop generic interfaces between user-level and VO-level resource brokering.

1.3.3. Grid Service Technologies

Our user-level resource monitoring, discovery, and brokering effort aims to deliver Grid services for functions that are potentially useful to Tier-2 users. The technologies employed to develop these services (Figure 3) are compliant with the evolving OGSA (Open Grid Service Architecture). We have prototyped a Grid-based problem solving environment – GISolve that is used to perform computationally intensive geographic information analysis (Wang, 2004). The implemented technologies and methods are generic, and can be adapted to support CMS data analysis for end users. Based on our experience in developing the interface between CMS data handling and analysis, we wish to work with Fermi Lab, and the Caltech Tier-2 Clarens[4] team as well as other Tier-2 centers, to build a Grid portal environment for supporting user-level CMS data handling and analysis.

2. UI Physics and CMS Activities

The University of Iowa High Energy Physics (HEP) and Nuclear Physics (NP) experimental and theory groups have seven HEP faculty and four NP faculty. The Physics Department has about 32 faculty members as well as 80 graduate students and 100 undergraduate physics majors. In HEP the major activities are the CMS-pp program (Y. Onel, E. McCliment, E. Norbeck), U. Mallik (BABAR) and C. R. Newsom (BTEV). The CMS physics program is supported by Y. Meurice, M.H. Reno and V. Rodgers (HEP Theory) and their graduate students.

In experimental NP, there is only one CMS activity, the CMS heavy ion program of E. Norbeck and N. George (Upper Iowa) and their graduate students. The NP group is involved in the design and construction of the ZDC as well as in the high level trigger for heavy ion reactions and physics analysis. They have also been testing prototypes of novel detectors for possible use in the ZDC and the CMS upgrade.

The CMS group of Y. Onel has presently about 22 members, however most of them are part-time on the CMS project. Four graduate students and four undergraduate students are 50%. Our three high-school teachers are full time during the summer but 10% during the school year. We keep two full time personnel at CERN. Our engineer Ianos Schmidt is responsible for the HF Wedge component assembly and integration of the HF with other future forward detectors like TOTEM, CASTOR and ZDC. He is also coordinator for fiber insertion and for HCAL cables and racks as well as HCAL radioactive source calibration hardware and electronics readout. Research Associate J.P. Merlo is responsible for the radiation damage studies for the HF detector components.

We are actively participating in physics simulation and analysis – HF and HE detector simulations for the LHC and SLHC, HCAL databases/calibration, photodetector (PMT and HPD), PMT-base circuit board, LED driver circuit board, and cables. We are simulating Higgs production via WW and ZZ fusion with forward jet tagging. We have already submitted a technical note to CMS (under review by the JETS/MET group) and are preparing a second note that includes effects from pile-up. We will maintain/upgrade what we have built for CMS as well as maintain "Data Base" and develop “Physics Analysis”, “VCR”, “Test Beam Analysis” and participate fully in “Installation-Commissioning”, and “Slice Tests Programs”.

The University of Iowa has a strong outreach program through Quarknet. We were one of the first Quarknet centers for CMS and have directed highly successful Quarknet Teacher Institutes in 2000, 2002 and 2004. We have about 30 high school teachers from Iowa in our program. In addition, since 2000 we have run Quarknet Summer research programs every summer with our three principal Quarknet teachers and selected high school students. We are also in close contact and doing collaborative work with the Iowa State University (Group W. Anderson) and several Iowa Small Private colleges like Upper Iowa University and Coe college. We are also doing collaborative work with other midwest universites like U. of Kansas and Kansas State University.

CMS HF project (1994-): We proposed quartz fiber calorimetry for CMS HF in January 1994 after prototyping the quartz fiber calorimetry using SSC GEM Closeout funds. There are now 6 US and 9 international institutions (15 in total) in the

[4] http://clarens.sourceforge.net/

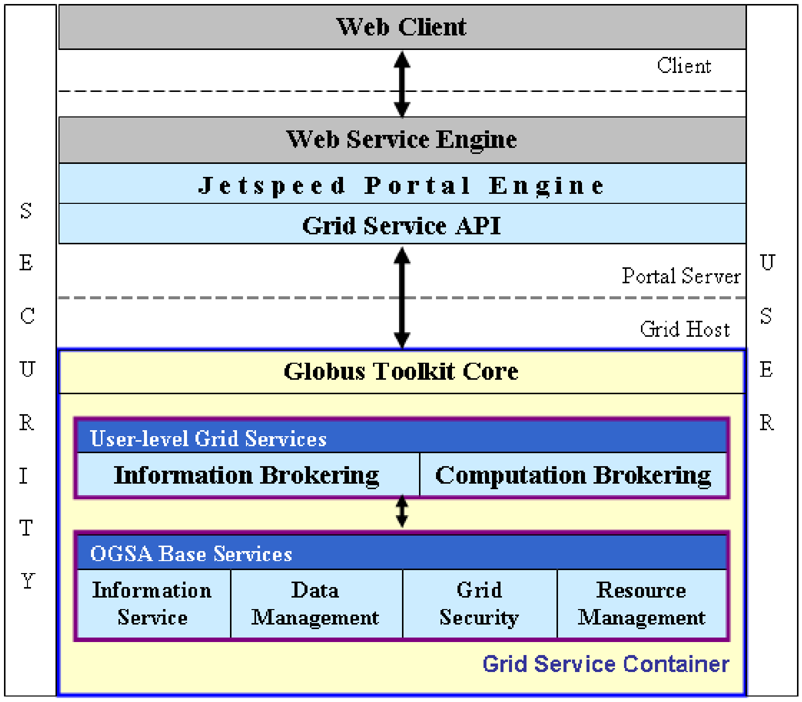


Figure 3. Grid service technologies

CMS HF group. We have taken a lot of responsibilities for the HCAL construction and have delivered all our tasks on time. The US CMS HF group at Iowa is responsible for:

1- HF detector prototypes:

i) Engineering design of prototypes, pre-production prototypes and manufacturing of the modules and components;

ii) Engineering design and manufacture of the Readout box for the pre-production modules;

iii) Engineering design and manufacture of the optical system for the pre-production modules;

iv) Engineering design of the HF calibration system (LASER and LED) and development of the source calibration systems for the pre-production modules; and

v) Production and engineering design of the HCAL LED drivers (HB, HE, HO and HF) and manufacture of prototypes.

2 - Selection and purchase of US quartz fibers and PMT's in addition to the responsibilities of procurement procedures, contracts, insurance, QC at manufacturer, delivery schedules and final delivery;

3 - Fiber radiation damage tests and studies at IOWA LIL/CERN facility and light-guide reflector material radiation damage tests at Iowa;

4 - Construction of the CMS-HF IOWA PMT test station facility and test/ quality control of the HF PMT's and maintenance of the web based database;

5 - Design and construction of HF light guides for the first two wedges and procurement of the light guide material for the remaining wedges;

6 - Design and construction of source distribution mechanics, including source tubing couplers and coupler pins at Iowa machine shop;

7 - Design and construction of HCAL LED drivers and related electronics;

8 - Construction of HF PMT Base boards (PCB) and QC/Testing;

9 - Design, construction and testing of HF on-line radiation monitoring system;

10 - HCAL Source Project- Readout electronics for driver system and mechanical upgrade.

11- Fiber insertion/optical assembly- Construction of a full scale wedge mock-up, development of the fiber processing and assembly, coordination and supervision of the fiber insertion and optical assembly; and

12- Cables and Services for HF- Design of service routing and evaluation of service requirements, comprehensive wiring diagram for all on-detector HF systems-including all services, data, electric and optical interfaces and specification of all cable types for HF services.

More information about our physics and CMS activities can be found in the following URL links: http://pion.physics.uiowa.edu/DOE/, http://pion.physics.uiowa.edu/hep/, http://pion.physics.uiowa.edu/DOE/photo2.htm, http://pion.physics.uiowa.edu/qnet/
http://www.physics.uiowa.edu/HCAL/.

3. Leveraging UI Resources for Tier-2

The UI CMS proof-of-concept Tier-2 center has been established by efficiently and effectively leveraging existing university resources as well as through continuing and significant university support. The initial funding of **$44K** (for personnel) and **$15K** (for DGT system) was provided by the UI Vice President for Research Office and ITS. ITS has also funded the technical director and several graduate students working on Tier-2 tasks and housed the systems of the center since the center was established. More importantly, the center has developed a proof-of-concept campus Grid – HawkGrid based on physics Grid technologies.

3.1. HawkGrid

The proof-of-concept HawkGrid operation center has been established to implement the following four major functions:

- Operating a viable GSI that will create and manage certificate authorities for dynamic VOs;
- Monitoring the health of HawkGrid computing resources and providing Grid software troubleshooting support;
- Providing education and consulting services for operating the Grid and "Gridifying" specific applications; and
- Conducting research on Grids by focusing on their applications to problem solving in various scientific and engineering domains.

The first two functions handle the infrastructure aspects of HawkGrid, while the last two specifically address the needs for enabling research and education using HawkGrid. HawkGrid is architected using VO concept and technologies that the physics Grid community uses as well. For example, two HawkGrid VOs (Figure 4) are shown having an overlapping resource: "rtgrid1.its.uiowa.edu," managed by Information Technology Services (ITS) (Armstrong *et al.*, 2004). The authentication polices for both VOs are based on the Grid Security Infrastructure (GSI) protocol (Butler *et al.*, 2000). The ITS-Stat-Geo VO was created to provide a Grid environment for a collaborative project between researchers in geography, statistics and ITS. This VO is comprised of two cluster resources ("rtgrid1.its.uiowa.edu" and "cluster0.its.uiowa.edu") managed by ITS and one cluster resource ("beowulf.stat.uiowa.edu") managed by the computer support group in the UI School of Mathematics. The Grid middleware configuration of the VO includes VDT (virtual data toolkit) as well as PBS and MPICH-G2.

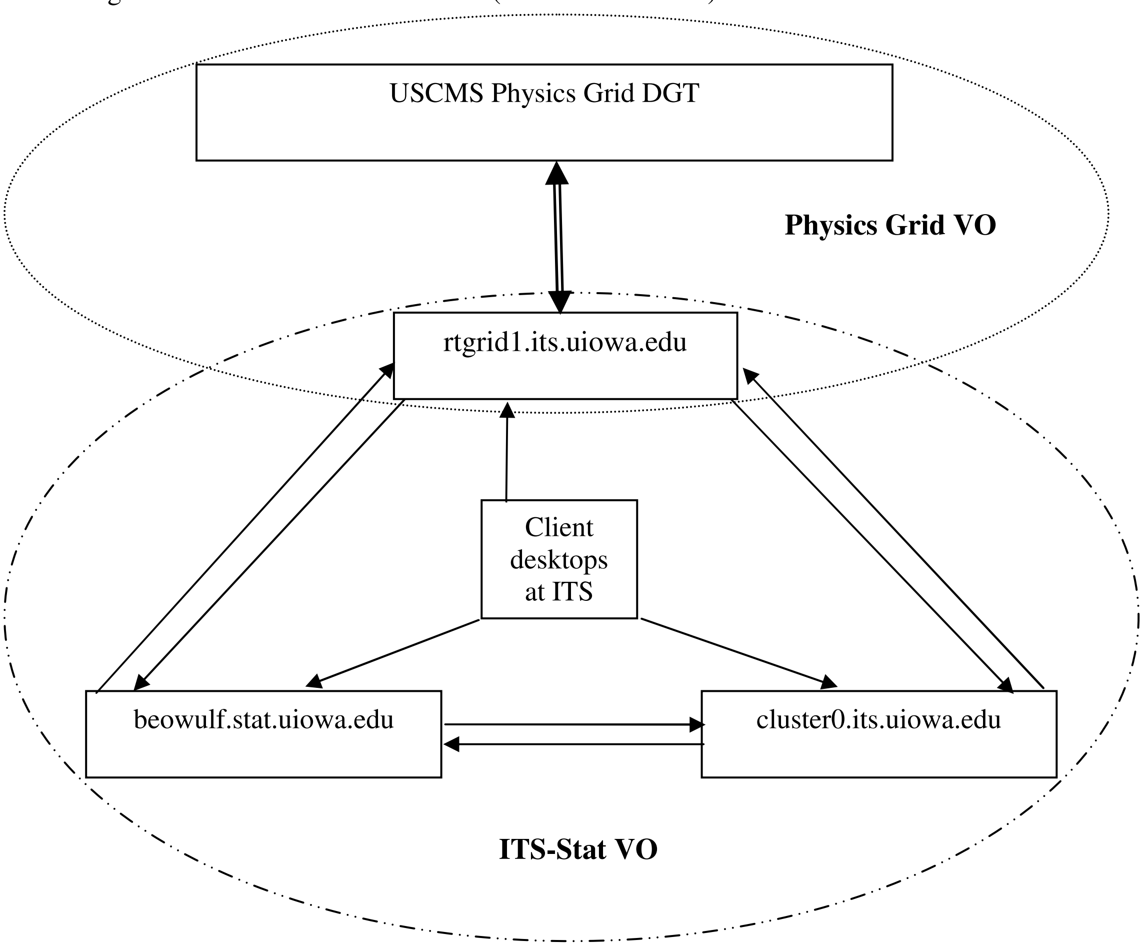


Figure 4. Two overlapping HawkGrid VOs

This configuration will be expanded to include several selected middleware tools from the TeraGrid software stack[5] and Open Science Grid deployment suite based on the needs of future Tier-2/Tier-2C users. HawkGrid will act as a Grid gateway for Tier-2 to opportunistically use the resources (see Section 4.c) made available from our Tier-2C academic units.

3.2. Staff Resources

Significant UI staff resources have been leveraged to investigate, develop, and support the proof-of-concept center functions (Table 1). A graduate student (Ransom Briggs) who works with the center technical director (Dr. Shaowen Wang) is funded as a 50% Graduate Research Assistant from the USCMS Software and Computing Project. UI will continue to provide significant university staff resources that are leveraged or matched to support the Tier-2 project.
However, the additional funding requested in this proposal must be provided to further support the UI Tier-2 program.

Table 1. Staff for the proof-of-concept Tier-2

Name	Affiliation	Tasks	%
Ugur Akgun (PD)	CMS	CMS Analysis Tools	25%
Ahmet Sedat Ayan (PD)	CMS	CMS Analysis	25%
Kevin Dolan (U)	CMS/EE	CMS Simulations	25%
Boyd M. Knosp (R)	ITS	Project Administration	10%
Yan Liu (G)	CS/ITS	Grid Scheduling & Grid-Enabled Databases	25%
Alexi Mestvirishvili (R)	CMS	CMS Data Analysis Support	25%
Derek D Monner (U)	CMS/CS	CMS Data Analysis Support	25%
Jun Ni (R, P)	ITS/CS	Grid Middleware and Database	10%
Ed Norbeck (P)	CMS	Overall Consultation	10%
Jonathan Olson (G)	CMS/EE	Grid Middleware, CMS Analysis Tools	25%
Yasar Onel (P)	CMS	Overall Director	20%
Anand Padmanabhan (G)	CS/ITS	Grid Resource Monitoring / Discovery	50%
Shaowen Wang (R)	ITS	Technical Director	50%

Note: G: Graduate Research Assistant; PD: Postdoc; P: Professor; R: Research Scientist; U: Undergraduate Research Assistant; CS: Computer Science; EE: Electrical Engineering; ITS: Information Technology Services.

4. Response to Selection Criteria

This section presents a comprehensive summary of the response to the selection criteria listed in the RFP.

a. Dedicated US CMS CPU and disk capacity expected to be achieved using Tier-2 program equipment funds (in SI2000 of CPU, Terabytes of storage).

   The floor space, reliable conditioned power, cooling available to our Tier-2 equipment are flexible to accommodate any installation and deployment plans. Consequently, this flexibility allows our Tier-2 to serve as a full-capacity (specified from the Caltech Tier-2 meeting) center in both CPU and storage. Though operated in a small scale for the three diversified systems that cover the cluster technologies used by the prototype CMS Tier-2 centers, our proof-of-concept Tier-2 has gained necessary system knowledge required to operate the full-capacity Tier-2 equipment. In addition, because our proposed Tier-2 will be operated by the university ITS, the staff of which have already had experience of operating data centers for the entire university. We will have straightforward access to the knowledge of security and networking. We are also willing to work with the Tier-1 and other Tier-2 centers to design our equipment deployment and operation plan that is consistently optimal to spend our Tier-2 equipment funding based on hardware cost and other factors.

b. Dedicated US CMS CPU and disk capacity expected to be available from local leveraged resources (based on alternate funding) (in SI2000 of CPU, Terabytes of storage).

[5] http://www.teragrid.org/userinfo/guide_software.html

Processors	Peak	Total Memory	Total Local Disk	Storage	Network	SI2000
22 Xeon @ 2.4 GHz	52.8 Gflops	14 GB	1760 GB	1TB SCSI RAID	1Gb/s to Internet2	19549.2
6 Athlon 2200+ @ 1.8 GHz	10.8 Gflops	4 GB	240 GB	0.5TB SCSI RAID	1Gb/s to Internet2	4311
10 PIII @ .65GHz	6.5 Gflops	2.5GB	200G	None	1Gb/s to Internet2	2600
Total				**1.5TB SCSI RAID**		26460.2

These three Linux clusters are currently used as the UI CMS proof-of-concept Tier-2 systems.

c. Expected useful annual integrated CPU capacity from leveraged non-dedicated resources (in SI2000-years).

Department / Experiment	Processors	Peak	Total Memory	Total Local Disk	Storage	Network	SI2000
CBCB-1	36 Athlon 2400 @ 2.0 GHz	72 Gflops	36GB	720 GB	2.5TB RAID	1Gb/s to Internet2	27.342
CBCB-2	32 PIII @ 500 MHz	16 Gflops	16 GB	9GB		1Gb/s to Internet2	6.4
IIHR	128 PIV @ 3.06 GHz	391 Gflops	256GB		8TB RAID	1Gb/s to Internet2	139.75
Radiology	64 (Athlon 3400+ 64bit) @ 2.20 GHz	140.8 Gflops			10TB RAID	1Gb/s to Internet2	85.354
ITS	32 PIII @ 1GHz	32 Gflops	16 GB	400 GB	.25TB RAID	1Gb/s to Internet2	12.8
Statistics	28 PIV @ 1.5 GHz	42 Gflops	14GB	257.6GB	0.5TB RAID	1Gb/s to Internet2	14.739
CGRER	68 (Athlon 3200+ 32bit) @ 2.20 GHz	149 Gflops			3TB RAID	1Gb/s to Internet2	70.992
Total					**24.25TB RAID**		357378.26

The table above shows the CPU capacity that is made available through HawkGrid to our Tier-2. Most of these Linux clusters are research-oriented equipment. They are likely to be upgraded from FY2005 through FY2009. But we cannot provide exact prediction for the growth of each system capacity. It might be appropriate to use the ratio of **10%** as the average annual increase for the capacity of these systems based on the past funding record for acquiring and upgrading these systems. The annually integrated capacity of these HawkGrid resources that is available to CMS will depend on the negotiated HawkGrid resource sharing policies.

d. The number of dedicated US CMS FTEs supporting Tier-2 operations, funded out of the Tier-2 program (in FTEs).

Affiliation	Name	Role	Contributions	%FTE
ITS	Shaowen Wang	Co-PI, Technical Director	Tier-2 Grid Support	25%
ITS	TBD	Grid Administration	System Support	100%

e. The number of dedicated FTEs supporting Tier-2 operations leveraged from other funding sources (in FTEs).

Affiliation	Name	Role	Contributions	%FTE
ITS	Shaowen Wang	Co-PI, Technical Director	Technical management and leadership	25%
Physics	Post Doctorate Researcher	CMS and Grid Software Researcher	Installing, configuring, and supporting CMS and Grid software, assist with Grid operation work	100%
Physics	Graduate Fellow	CMS and Grid Software Researcher	Helping with testing and developing Grid-based CMS data handling and analysis software	50%
Physics	Yasar Onel	The PI	Raising Tier-2 funding, providing scientific leadership, and coordinating education and outreach activities	25%
Physics	Edwin Norbeck	CMS Data Analysis Consultant	Assisting with configuration, testing and utilization of CMS software suite, promote the use of Tier-2 resources	25%
Physics	Alexi Mestvirishvili	CMS Data Analysis Research Scientist	Assisting with testing of CMS software	50%

f. Expected useful FTE effort supporting Tier-2 operations but not funded by or dedicated to US CMS (in FTE). Please give indications of how and in what function these FTE will contribute.

Department/group	Name	Role	Contributions	%FTE
ITS Research Services	Boyd M. Knosp	Group Leader	Education and outreach, configuring ITS facilities for OSG, helping with managing Tier-2 operations	25%
ITS Research Services	Jun Ni	Research Scientist	Education and Outreach, configuring ITS facilities for the grid, Grid research activities	10%
ITS Systems and Platform	Systems staff	Systems engineer	Assist with system configuration and facilities support	10%
ITS Telecommunications and Networking	Jay Ford	Network Engineer	Assist with configuring and testing the network	10%
IIHR Hydroscience and Engineering	Mark Wilson	Data Systems Coordinator	Assist with configuring and testing the grid and providing access to IIHR resources	10%
CGRER	Jereme Moen	Systems Engineer	Assist with configuring and testing the grid and providing access to CGRER resources	10%
CBCB	Systems Staff	Systems Engineer	Assist with configuring and testing the grid and providing access to CBCB resources	10%

g. Wide area network connectivity, and expected costs and funding sources to make this bandwidth available to the Tier-2 center, in case additional equipment is required.

Network organizations: The UI is a founding member institution of Internet2. As a member institution of the CIC (Committee on Institutional Cooperation), we have a shared membership in the National LambdaRail, Inc. We are also a founding member of a regional network initiative that includes the University of Wisconsin, the University of Minnesota, and Iowa Statue University called the Northern Tier Network. This regional group is evaluating a dark fiber ring.

We have a 1Gb/s connection to Internet2 that is separate from our connection to the commodity Internet. This connection goes to the Abilene hub in Chicago. We also have a redundant 1Gb/s Internet 2 connection via Iowa State and the Great Plains network.

We are planning to upgrade our current network connection using Wavelength Division Multiplexing (WDM) dark fiber networking with the goal of having it installed by calendar year 2007. Two of the scenarios that we are pursuing towards implementing this plan are to:

1. work with the Northern Tier Network to establish a WDM regional network that increases our network bandwidth to 10 Gb/s or higher.
2. identify and purchase dark fiber between Iowa City and Chicago and increase our network bandwidth to 2.5 Gb/s or higher.

Costs for this network upgrade are anticipated to range from **$1 to $1.5Million** and yearly costs will be around **$100,000**. Funding will be provided by The University of Iowa.

5. Operations Model

The proposed Tier-2 center will provide 8am-5pm, Monday through Friday attended operation services. Off hours (including weekends) call services as well as email list support are provided. This operation model is similar to the operation model adopted to support our campus high performance computing systems by ITS research services. Given our experience of supporting these systems and working on the proof-of-concept systems, we are confident that we will provide reliable site services such as administration and registration services. Security and incidence response services will be provided by Tier-2 staff who will interact with ITS System & Platform Administration (SPA[6]) as well the Tier-1 and other Tier-2s to deliver instant recover solutions. Also, we are flexible to work with other Tier-2 centers to provide other types of operation models as required by dynamic user requests and status of other Tier-2 resources.

II. The University of Iowa Tier-2C Program: Empowering Cyberinfrastructure Application Communities through the Open Science Grid

1. Introduction

The University of Iowa (UI) has a strong interest in developing cyberinfrastructure communities through the use of advanced Grid computing capabilities. Nine domains (see Appendix) will contribute to the Tier-2C program and advance pertinent community-specific knowledge environments by taking advantage of existing Open Science Grid (OSG) capabilities. Meanwhile, an ultimate goal of these domains is to contribute experience of adapting applications to future OSG development through our proposed Tier-2 center. Three categories of OSG applications have been identified from these domains (see Appendix A).

2. Tier-2C Applications

Our Tier-2C applications are designed based on the experience and in-depth knowledge of our team members in Grid and high performance computing applications and their leadership roles in developing regional and national cyberinfrastructure initiatives.

Atmospheric science: Assimilating immensely complex and large earth-science data into Chemical Transport Models (CTM) is recognized as essential in weather/climate analysis and forecast activities. However, data assimilation poses a significant computational demand for high CPU speed, large storage capacity, and high performance of inter-processor communications that are often not available at a single computer level. We plan to develop Grid-based algorithms for efficient data assimilation in CTMs based on a communication library that was developed for parallel computing of data assimilation (Miehe *et al.*, 2002). Grid-based PSE for CTMs will be studied to integrate general computational tools and associated software that exploit Grid computing to assimilate atmospheric chemical and optical measurements into CTMs. By doing so, we also wish to develop collaborative visualization environments in which users need not be experts in Grid computing, or adjoint modeling and optimization theory to use CTMs with data assimilation capabilities that takes advantage of OSG capabilities. Our application development work is placed within the context of developing national Cyberinfrastructure for Research and Development in the Atmospheric Sciences (CyRDAS) for which Dr. Gregory R. Carmichael serves as a steering committee member.

Bioinformatics: The field of bioinformatics has, for many years, been a significant consumer of computational resources. Some of the computationally intensive applications used by the Center for Bioinformatics and Computational Biology at UI include: sequence alignment, sequence analysis/annotation, statistical analysis of microarrays, phylogenetic analysis via sequence alignment and gene clustering hidden markov modeling. These computational needs are expanding,

---

[6] http://www.its.uiowa.edu/spa/

particularly for sequence alignment, an application that is a foundation of the field. Sequence alignment, algorithmically, is largely considered solved, but the accumulation of sequences in databases is increasing exponentially leading to slower alignments and storage problems.

While we currently have a computational cluster that partially addresses these problems, this will not suffice in the future, and the current scheduling methods we are using now will not work well for the applications we are envisioning. Through the use of OSG technologies, we foresee several changes to our current processes: making the cluster more usable to a broader subset of researchers, the mirroring of, or intelligent access to, many gigabyte datasets on every computational node in an intelligent way, and providing high bandwidth access to unique currated datasets from around the country. Ideally, researchers with a biological rather than a computational background could then more directly use these resources. Currently the use of these clusters is confined to a limited subset of researchers.

Biomedical Imaging: X-ray CT is an important non-invasive medical imaging technique that plays an essential role in modern medicine. High-end parallel computers may support the development of feasible solutions to computationally intensive CT image reconstruction algorithms (e.g., iterative image reconstruction). However, current approaches are often too expensive to be practically used and maintained. Grid computing is promising because it enables computing resources to be dynamically aggregated in a scalable way. In an earlier study, a peer-to-peer system was developed to improve the performance of an iterative image reconstruction algorithm by exploiting its parallelism (Ni *et al.*, 2004). We expect that OSG will become more decentralized and take advantage of peer-to-peer technologies. We plan to further study resource scheduling and load balancing techniques that take advantage of both peer-to-peer and Grid computing capabilities.

OSG technologies are also potentially useful to develop finite-element reconstruction algorithms used in our bioluminescence tomography (BLT) (Wang *et al.*, 2004), and dynamic elastography (DE) (Liu *et al.*, 2004) projects. In the BLT case, finite-element mesh generation involves intensive computation. To capture the anatomic components in small animals, such as heart and lungs, hundreds of high-resolution (512 × 512) CT slices are needed, which yield a few hundred million tetrahedral elements (Cong *et al.*, 2004). These elements need to be smoothed and optimized before they are used for BLT reconstruction. Such computation can be divided into a large number of independent sub-tasks, for which Grid computing is well adapted. In the DE case, finite-element mesh creation is dynamic, and becomes even more challenging than BLT.

Computational Science: For matrix assembly in numerical solvers, appropriate data structures are important and resulting systems are solved efficiently with iterative methods such as Multigrid/Multilevel Algorithms. Data distribution in parallel computing is tightly connected to a graph representation of the task, and splitting this graph into subgraphs (representing substasks), minimizes dependence between subtasks. Graph partitioning techniques include continuous optimization and eigensolvers. We anticipate that these models will be useful for task distribution in Grid computing environments.

Computer Science: Applications-oriented research in distributed systems focuses on solutions to combinatorial problems that arise in diverse fields, ranging from, e.g., automated theorem proving, satisfiability checking and graph coloring problems to genetic linkage analysis, protein structure prediction, and winner determination in combinatorial auctions. Many of these problems can be cast as search problems, and then solved either by heuristic means or more exhaustive search techniques (e.g., B&B, simulated annealing, genetic search, and so on), which may or may not guarantee the optimality of solutions. Note that such algorithms are difficult to parallelize effectively, and, unlike most Grid applications, generally do not involve an overwhelming amount of data. Thus, we tend to hit computation bounds rather than communication bounds. One important focus of departmental research is on the design of appropriate middleware and infrastructure components to support algorithmic development of solvers for these kinds of problems (Segre, 2002).

Geographical Information Science: Geographic information analysis methods often require the use of heuristic search, simulation, optimization, and statistical methods that are computationally intensive, particularly when they are applied to large, realistic problems (Densham and Armstrong, 1998). We wish to improve our domain decomposition and task scheduling methods that were developed for Grid computing of spatial interpolation, a widely used geographic information analysis function (Wang and Armstrong, 2003). It is feasible for these methods to be designed as generic application-level Grid services that support Grid computing of geographic information analysis (Wang, 2004). These services will be integrated with other generic Grid services that are compliant with the Open Grid Service Architecture to create a problem solving environment – named GISolve. GISolve will be developed to take advantage of Virtual Data Toolkit capabilities that continue to evolve with future OSG development.

Environmental Sciences and Engineering: A Center for Environmental CyberEngineering (CECE): Sensing, Modeling, and Systems Analysis is being planned to establish a collaborative engineering analysis network, using high performance tools and Grid infrastructure, to transform our understanding of environmental change in human-dominated systems, and to accelerate development of cyberengineering and the sensor industry. Dr. Jerald L. Schnoor was a workshop planner and participant in the Cyberinfrastructure Research and Education Workshop held at the National Center for Atmospheric Research in October – November, 2002. Dr. Schnoor is currently leading a multi-institution effort to design the CECE. We wish to experiment with OSG technologies to analyze critical national environmental problems while developing a whole new industrial technological system for sensing, networking, and modeling the environment.

Management Science: Certain large-scale optimization problems are well posed for implementation on computational Grids. Examples include branch-and-bound (B&B) methods for integer and mixed-integer programming, and algorithms for stochastic optimization that utilize Monte-Carlo sampling. In 2000 the Grid-based implementation of a nonlinear B&B algorithm solved the "nug30" Quadratic Assignment Problem, an open challenge problem in discrete optimization posed in 1968 (Anstreicher, 2002). The solution of nug30, utilizing an average of 650 CPUs for over 7 days, was a significant early success in Grid computation.

Computational Grids are also well suited for global optimization problems involving highly nonlinear functions, such as those occurring in geometric optimization problems. Protein folding is a well-known example with very significant applications. Similar examples arise in the analysis of other geometric problems in three dimensions, such as the Kepler conjecture. The proof of the Kepler conjecture by T. Hales and S. Ferguson, first announced in 1998, required the solution of over 100,000 linear and mixed-integer programming problems. The proof is currently considered "uncheckable" due to the enormous amount of computation (performed by Hales and co-workers over a several-year period). Reproducible verification of the Hales-Ferguson proof will likely require a large-scale OSG-based implementation to permit more efficient execution of the required computations.

Statistical Computing: Strategies are being investigated to use Grid computing to enable statistical analysis of massive spatial and spatiotemporal datasets. For the Gi* statistic – a measure of spatial clustering proposed in Getis and Ord (1992), and further elucidated in Ord and Getis (1995) – application-specific middleware for efficient Grid-based computation (Armstrong, Cowles, and Wang, 2004, Wang, 2004) has been developed based on HawkGrid. A pilot study is being conducted to examine Grid-based computational methods for fitting classical and Bayesian geostatistical models. Initial development is being conducted on the HawkGrid, and final testing will be done on the TeraGrid. After completion of the pilot study, we will develop Grid-based methods for fitting more complex Bayesian spatial and spatiotemporal models, and will use them to analyze environmental health data. Participation in Tier-2C will provide the computing environment for carrying out this work.

An infrastructure is being developed for parallel computation from within the high level statistical data analysis language R (Rossini *et al.*, 2003). The initial interface is designed to handle problems that can be decomposed into a sequence of scatter-compute-gather steps on a computational cluster. This framework will be extended to allow more sophisticated communication and to take advantage of OSG resources. Extension to the Grid framework will require addressing Grid-specific issues, such as managing authentication, fault tolerance, and fault recovery from a high level language, and dealing with variations in computational and communication performance among computational nodes and over time. The current prototype implementation has been used successfully in bioinformatics, geostatistics, and financial analysis; these and other areas will benefit from extensions to a Grid framework.

3. HawkGrid

HawkGrid will be used as an intra-campus Grid computing environment to support the development of the Tier-2C applications described above.

3.1. Application Development Sharing

A production HawkGrid operation center will be established to provide an interface between Tier-2 and our campus computing resources (see Section 4.c from part I) organized as dynamic virtual organizations. More specifically, this interface will be implemented by applying or advancing OSG-compatible technologies to enable application development sharing facilitated by an application development team that consists of multi-disciplinary personnel. The areas of application development will focus on developing application- and user-level Grid computing capabilities. The development of these capabilities is strengthened by the long-term Grid computing Research and Development roadmap described previously as part of the Tier-2 proposal.

3.2. Fostering Interdisciplinary Research among Tier-2C Applications

The selection and design of our Tier-2C applications described above is designed to foster collaborative problem solving and interdisciplinary research. Several preliminary studies of the Tier-2C applications have involved interdisciplinary research among our Tier-2 team members. For example, researchers from Statistics (Cowles, Yan) and GIScience (Armstrong, Wang) have been working together on using Bayesian methods to analyze complex geographic information; researchers from Atmospheric Science (Carmichael *et al.*), Environmental Sciences and Engineering (Schnoor *et al.*), and GIScience (Armstrong) have collaborated to investigate Environmental CyberEngineering; researchers from Biomedical Imaging (Wang *et al.*) and Computer/Computational Sciences (Ni, Wang) have developed a peer-to-peer system to implement an efficient image reconstruction algorithm. Through the proposed Tier-2C program, we have developed an understanding that each individual application will benefit from the experience of other interdisciplinary applications by sharing a common OSG-compatible development environment – HawkGrid.

III. Education and Outreach

The education and outreach program associated with the Tier-2/Tier-2C will support a collaborative learning community engaging traditional and non-traditional students in our studies of applications in the cyberinfrastructure frontier. We will develop a HawkGrid learning environment in which students will become empowered to design and implement OSG application projects and/or help develop the learning environment itself under professorial supervision. A key aspect of this activity is that faculty, students, and staff will be engaged closely in the educational activities. A key challenge in the utilization of cyberinfrastructure is the development of human-resources. At The University of Iowa the expertise is dispersed between the faculty, students and staff. The design of our Tier-2/Tier-2C proposal recognizes this and we have outlined a structure (see V. Management Plan) in which these groups can come together in a collaborative environment to conduct important research. The educational component will consist of research teams involving staff, students and faculty. In addition, we plan to conduct seminars and mini-workshops, lead by students, staff, and faculty, where in-house expertise and best-practices will be shared. We also envision that new undergraduate and graduate level courses will be developed (see the support letter from our Computer Science Department).

We will deploy OSG technologies and learning environment tools into high schools and university classrooms to disseminate cyberinfrastructure knowledge, and explore how the tools adapt to different school communities. Our education and outreach team will continue to leverage the three existing education and outreach programs: Iowa QuarkNet, Science Education Center of School of Education, and the Studio of Academic Technologies for Education.

IV. Broader Impact

The proposed Tier-2/Tier-2C activities will broadly impact research and education activities at The University of Iowa, the Midwest region, and the entire county. The undergraduate and graduate student training on campus will benefit directly by the enhanced exposure to the interdisciplinary elements of cyberinfrastructure in general, and Grid computing in particular. Experienced and knowledgeable faculty and staff from **9 disciplines** within **22 academic units** who have significant track record in advancing computer and computational sciences will participate in and contribute to the Tier-2C development. The enhanced institutional capabilities will also broadly impact the educational and research activities in units not directly identified in this proposal. Especially, through the GRid Application Funding Initiative (GRAFI) funded by the UI Office of Vice President for Research Internal Funding Initiative (see the support letter from the UI Office of Vice President for Research), Tier-2/Tier-2C knowledge can be advanced by all disciplines available at The University of Iowa. The experiences gained through the proposed Tier-2/Tier-2C project, will be shared, and used to help identify how these tools can be used in other research and education projects, including those arising in the arts and humanities.

V. Management Plan

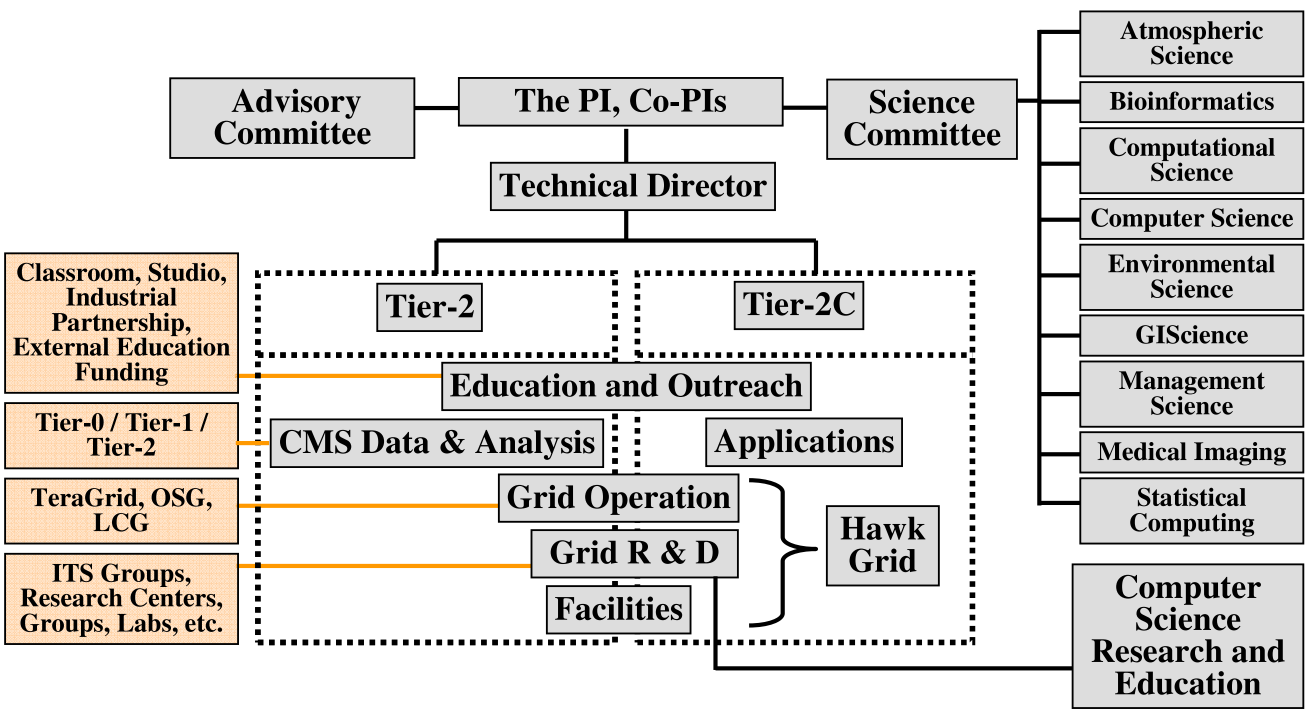
The University of Iowa Tier-2/Tier-2C project will provide dependable Grid-based cyberinfrastructure services to the CMS collaboration and will implement a broad and coherent scientific program that takes advantage of Open Science Grid (OSG) capabilities. The success of the project will be assured by a strongly committed collaboration among Information Technology Services (ITS), the CMS Data & Analysis Team, and the UI computer and computational science research and education community. In this collaboration, ITS plays a bridging role by bringing together people who will innovatively develop and use evolving OSG capabilities. This unique role played by ITS is carried out by the Grid Research and educatiOn group at Iowa (GROW).

GROW has represented The University of Iowa during the past two years in several large-scale Physics Grid projects such as iVDGL, DAWN, and GECSR as well as the largest international Grid testbeds – LCG0 and LCG2. GROW has been used as a management framework to make The University of Iowa a regional Grid computing center for USCMS collaboration. During this period of time, GROW has also made significant progress toward the establishment of a consortium that allows Grid computing professionals and domain scientists to work together on computationally challenging problems as demonstrated by our publications and TeraGrid awards. In our proposed Tier-2/Tier-2C project, GROW will continue to be used as a management framework to not only deliver reliable and robust cyberinfrastructure services to USCMS collaboration, but also to facilitate the creation of cyberinfrastructure at the university level to foster collaborative problem solving and interdisciplinary research among application scientists and computer/computational scientists.

The PI, Co-PIs and other senior personnel associated with the project will continue the success of GROW with their extensive experience in the management and execution of large collaborative projects. Professor Yasar Onel, the PI of the proposed Tier-2/Tier-2C project, is the USCMS HF coordinator. Professor Onel has been a GROW advisory committee member since GROW was founded. He will coordinate the overall management of the project, and will also be responsible for Tier-2 data analysis services. During the past three years, Professor Onel has been collaborating with Dr. Shaowen Wang in all the GROW Physics Grid projects. Dr. Wang is a Co-PI and the **Technical Director** of both the Tier-2 and Tier-2C parts of the project. Dr. Wang is also the technical director of The University of Iowa proof-of-concept Tier-2 center. He is the PI of the campus Grid project – HawkGrid that will be used to integrate Tier-2 and Tier-2C as well as other university resources. Dr. Wang will be responsible for technical aspects of the project including facilities, operation, Grid research and development, and education and outreach. Professor Gregory R. Carmichael, another Co-PI of the project, is an international leader in the development of emissions inventories for natural and pollutant substances and of chemical transport models at scales ranging from local to global. He has served as a steering committee member for the national Cyberinfrastructure Research and Development in the Atmospheric Sciences (CyRDAS). Professor Carmichael is the PI of a current medium NSF ITR project that investigates "the development of a general computational framework for the optimal integration of atmospheric chemical transport models and measurements using adjoints." Professor Carmichael is an active instructor and advisor, having supervised 29 M.S. and 26 Ph.D. students. He is also a co-director of the Center for Global and Regional Environmental Research and Associate Dean for Research of the College of Engineering at The University of Iowa. Professor Carmichael will provide leadership in the strategic development of the project as well as guidance and advice for developing OSG-based atmospheric scientific applications. Professor Ge Wang, another project Co-PI, is an IEEE fellow in the area of medical imaging. Professor Ge Wang will be responsible for facilitating a coherent Tier-2C scientific program. Professor Wang has worked with Dr. Shaowen Wang, among others, to pioneer the development of a peer-to-peer distributed system that implements a computationally efficient image reconstruction algorithm. His research is well supported by the National Institute of Health. Professor Wang is the director of The University of Iowa High-Performance Computing Lab for Medical Imaging, and he is also actively developing a cyberinfrastructure research program for medical imaging at The University of Iowa and beyond.

The **GROW Management Structure** shown in the figure below will tightly connect the PI and Co-PIs with other senior personnel. The PI and Co-PIs will serve as the primary point of contact for USCMS/NSF and external relations, and will provide overall scientific leadership. The PI and Co-PIs will work closely with other senior personnel through several committees/teams that interact with each other for effective project execution.

The **Advisory Committee** co-chaired by Professor Gregory R. Carmichael (Co-PI) and Professor Thomas L. Casavant will provide general oversight for all aspects of the project as well as input on project vision and strategic planning, and will assist in developing opportunities for additional project funding. The Advisory Committee consists of selected central university administrators, deans of colleges involved, department chairs, and directors of research centers involved. Committee members include Professor Kurt M. Anstreicher, Professor Marc P. Armstrong, Professor Thomas F. Boggess Jr., Chief Information Technology Officer and Assistant Vice President for Research Steve Fleagle, Professor Virendra C. Patel, Professor Jerald L. Schnoor, and Professor Luke Tierney. The Advisory Committee will regularly meet with the PI and Co-PIs of the project with Professor Gregory R. Carmichael as the representative of the PI and Co-PIs.



The **Science Committee** co-chaired by Professor Kurt M. Anstreicher and Professor Ge Wang will be responsible for advising Tier-2 scientific activities and will help foster inter- and multi-disciplinary projects. This committee with representation from nine scientific and engineering domains, will review university internal proposals for using computational resources provided by the project to broaden the impact of Tier-2C on campus. The Science Committee will interact directly with the PI and Co-PIs and provide input to advise the Technical Director to execute the project and meet scientific goals.

Science Committee members include Professor Marc P. Armstrong, Professor Gregory R. Carmichael, Professor Yasar Onel, Professor Jerald L. Schnoor, Professor Alberto M. Segre, and Professor Luke Tierney.

Several **Technical Teams** have been organized to carry out the technical aspects of the Tier-2/Tier-2C project. These teams include the **Applications Team** co-chaired by Professor Marc P. Armstrong and Professor Mary Kathryn Cowles, the **CMS Data & Analysis Team** co-chaired by Professor Edwin Norbeck and a university-funded Tier-2 postdoc researcher, an **Education and Outreach Team** co-chaired by Boyd M. Knosp and Professor Yasar Onel, a **Facilities Team** co-chaired by Boyd M. Knosp and Mark A. Wilson, a **Grid Operation Team** chaired by a Tier-2 funded physics Grid staff member, and **a Grid Research and Development Team** co-chaired by Dr. Jun Ni and Professor Suely P. Oliveira. The Applications Team, Grid Operation Team, and Facilities Team will be integrated with pertinent leveraged resources of people and computation from HawkGrid. The tasks of these technical teams will be designed by the **Technical Director** with assistance from the team co-chairs. The **Technical Director – Dr. Shaowen Wang will devote half of his time to management and technical operation of Tier-2 and a quarter of his time to accomplishing Tier-2C technical goals.**

Members of the **Application Team** will include faculty who will receive a "Cyberinfrastructure Scholarship", research scientists, and graduate students whose funding has been leveraged from throughout The University of Iowa. The goals of this team are to develop and utilize OSG-based scientific applications, and to create a set of cross-disciplinary Grid computing tools to achieve specific scientific purposes. This team will drive the development of cyberinfrastructure for the graduate curriculum at The University of Iowa by working with the Education and Outreach team, and will provide case studies and documented best practices for cyberinfrastructure advancement and education.

The **CMS Data & Analysis Team** is comprised of a Tier-2 funded Physics Grid software specialist, CMS faculty, research associates, postdoc, and graduate students to provide CMS Tier-2 data analysis services. This team will work closely with the Application Team to share common practices and to leverage tools and experiences. This team will share experience with the Tier-0, Tier-1, and other Tier-2s regarding data handling and analysis services.

The **Education and Outreach Team** is responsible for developing and coordinating teaching curriculum, workshops, events, and communication activities. The team is made up of CMS faculty, research scientists, and a Web-designer to help design and coordinate the content management for the project web site. The project web site, an intranet site for sharing documents and organizing teams and committees, provides a shared location for depositing documents such as project proposals, reports, reviews, educational materials and software archives. The Education and Outreach Team interacts with the Applications Team and the CMS Data & Analysis Team to develop methods for training new scientists and students in Grid computing.

The **Grid Operation Team** consists of system support staff, system coordinators, and graduate students. The tasks of this team are to establish resource sharing procedures and policies as well as to manage a Grid Operation Center. The team has had extensive experience with securely configuring and managing OSG middleware while working with the Tier-0, Tier-1, and prototype Tier-2 centers and supporting the development of TeraGrid applications of campus researchers. The team will build on this experience, and leverage the HawkGrid Grid Operation Center resources to serve Tier-2 functions and support Tier-2C application development.

The **Grid Research and Development Team** will help faculty, research scientists, and students investigate, advance, and applyOSG technologies. This team studies Grid computing technologies in both empirical and theoretical ways. The knowledge gained from this team will advise the Grid Operation Team as they deploy reliable and robust OSG technologies. At the same time, this team will work closely with the Application Team to support application development.

The **Facilities Team** is composed of representatives from all the involved academic units. This team will allocate and install computing resources that include external and internal network connections.

Dr. Jun Ni, Professor Suely P. Oliveira, Professor Alberto M. Segre, Dr. Shaowen Wang, and their graduate and undergraduate students will form a **Computer Science Research and Education Committee** chaired by Professor Segre. This committee will carry out computer science research and education activities that are relevant to OSG and future cyberinfrastructure development through conducting specific research projects as well as regular computer science courses and seminars.

References

(The names of GROW members are highlighted.)

Anstreicher, K. M., Brixius, N., Goux, J.-P., and Linderoth, J., 2002, Solving large quadratic assignment problems on computational grids, *Mathematical Programming,* **91**, 563-588.

Armstrong, M. P., Cowles, M. K., and Wang, S., 2005, Using a computational Grid for geographic information analysis: a reconnaissance. *Professional Geographers*, forthcoming.

Cong, A., Liu, Y., Kumar, Y., D., Cong W.X., and **Wang, G.**, 2004, Geometrical modeling using multiregional marching tetrahedra for bioluminescence tomography. *Proc. SPIE (Medical Imaging) M105*, submitted.

Densham, P. J., and **Armstrong, M. P.**, 1998, Spatial analysis. In: *Parallel Processing Algorithms for GIS*, edited by Healey, R., Dowers, S., Gittings, B., and Mineter, M., Taylor and Francis, Bristol, PA.

Getis, A., and Ord, J. K., 1992, The analysis of spatial association by use of distance statistics. *Geographical Analysis*, **24**: 189-206.

Miehe, P., Sandu, A., **Carmichael, G. R.**, Tang, Y., and Daescu, D., 2002, A communication library for the parallelization of air quality models on structured grids. *Atmospheric Environment*, **36** (24): 3917-3930.

Liu, Y., **Wang, G.**, and Sun, L. Z., 2004, Tomography-based 3-D anisotropic elastography from boundary measurements II: Phantom Simulations. *IEEE Transactions on Medical Imaging*, submitted.

Ni, J., **He, T.**, Li, X., **Wang, S.**, **Wang, G.**, 2004, EM medical image reconstruction in a peer-to-peer systems. *Proceedings of Lecture Notes in Computer Science*, Volume **3251**, pp. 495-501, 2004.

Ord, J. K., and Getis, A., 1995, Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis*, **27** (4): 286-306.

Padmanabhan, A., **Wang, S.**, Ghosh, S., and **Briggs, R.**, 2004, An asynchronous peer-to-peer approach to Grid resource discovery using autonomous grouping. Submitted to *CCGrid05*.

Rossini, A. J., **Tierney, L.**, and Li, N., 2003, Simple parallel statistical computing in R. Tentatively accepted by *Journal of Computational and Graphical Statistics*.

Segre, A.M., Forman, S., Resta, G., and Wildenberg, A., 2002, Nagging: a scalable, fault-tolerant, distributed search paradigm. Artificial Intelligence, **140:** 71-106.

Wang, G., Li, Y., and Jiang, M., 2004, Uniqueness theorems in bioluminescence tomography, *Med. Phys*, **31**:2289-2299.

Wang, S., 2004, Grid-based geo-middleware for geographic analysis: theory, method, implementation, and evaluation. Ph.D. Thesis, The University of Iowa, Iowa City, IA, U.S.A.

Wang, S., Padmanabhan, A., Liu, Y., Briggs, R., Ni, J., He, T., Knosp, B. M., Onel Y., 2003, A multi-agent system framework for end-user level Grid monitoring using geographical information systems (MAGGIS): architecture and implementation. *Lecture Notes in Computer Science*, Volume **3032**, pp. 536-543.

Wang, S., and Armstrong, M. P., 2003, A quadtree approach to domain decomposition for spatial interpolation in Grid computing environments. *Parallel Computing,* **29** (10): 1481-1504.

Wang, S., Armstrong, M. P., and Bennett, D. A., 2002, Conceptual basics of middleware design to support grid computing of geographic information. In: *Proceedings of 2nd International Conference on Geographic Information Science*, September, 25 - 28, Boulder, CO, USA.

Appendix A. Categories of Tier-2C Applications.

	Simulation & Analysis	Problem Solving Environments (PSE)	Collaboratory
Atmospheric Science	Data assimilation in Chemical Transportation Models (CTM)	General PSE to facilitate the construction of discrete adjoints in CTMs using Grid computing resources	Cyberinfrastructure for Research and Development in the Atmospheric Sciences (CyRDAS)
Biomedical Imaging	Image reconstruction algorithms	________	________
Bioinformatics	Distributed data storage/access and sequence alignment	Grid-enabled bioinformatics computation	cancer Biomedical Informatics Grid (caBIG)
Computational Science	Graph partitioning methods	________	________
Computer Science	Distributed search	________	________
Geographical Information Science	Grid computing of geographic information analysis	GISolve for geographic information analysis	________
Environmental Sciences and Engineering	________	_______	Center for Environmental CyberEngineering (CECE): Sensing, Modeling, and Systems Analysis
Management Science	Large-scale optimization problems, global optimization problems involving highly nonlinear functions	_______	_______
Statistical Computing	Bayesian geostatistical models	Tools for accessing Grid resources from the high level data analysis language R	________