

## A. OBJECTIVE OF THE PROJECT

The Grid, as a well-defined hardware, software, research, education and outreach initiative, was first proposed in the 1990's [1], and has been pursued by researchers, institutions and funding agencies around the world since then, as a core enterprise of the discipline of High Performance Computing (HPC). One of the leading edge environments for Grid computing is the National Science Foundation's TeraGrid facility [2, 3], which includes substantial support for the acquisition, deployment and administration of Linux clusters that are based on Intel's IA-64 (Itanium) processor [4]. As the Teragrid comes online, the role of Itanium clusters in Grid computing is expected to increase substantially. Therefore, the goal of this project is to acquire, install, deploy and administer an Itanium cluster for Grid computing, as well as relevant support resources such as storage, software and administration, to be located at the University of Oklahoma (OU) and maintained by the OU Supercomputing Center for Education & Research (OSCER) [5]. The proposed acquisition will: (a) improve access to and increase use of cutting edge computational systems and education by scientists, engineers, graduate students and undergraduate students; and (b) enable a multidisciplinary team to create a well-equipped learning environment for integrating research and education.

Specifically, under this project, using a combination of NSF funds and OU cost share funds, OSCER plans to acquire: a Linux cluster consisting of a substantial number of Itanium processors, a large amount of memory, a high performance interconnect and substantial disk; considerable external storage, both high performance disk and tape; software necessary and appropriate for the Itanium cluster's planned user community (compilers and debuggers); significant system administrator time.

The research efforts that expect to employ these resources include multidisciplinary and multi-institution endeavors covering a wide variety of science and engineering topics: meteorology (two projects); earthquake engineering; high energy physics; multiscale flow through porous media for hydrocarbon reservoir simulation; aerospace design. In addition, the Itanium cluster will be used to teach HPC and Grid computing.

## B. RESEARCH AND EDUCATION ACTIVITIES

The Itanium cluster to be purchased under this project will be used primarily for Grid computing. Currently, OU research teams are pursuing six projects employing or expecting to employ Grid computing, three of which are already underway, two of which have proposals or preproposals pending, and one of which is in the proposal development stage. The resources will also be used to advance an innovative HPC education initiative, now underway. The initiatives are:

- Modeling Environment for Atmospheric Discovery (MEAD)
- Linked Environments for Atmospheric Discovery (LEAD)
- Data Grid Tools for High Energy Physics (HEP)
- George E. Brown, Jr. Network for Earthquake Engineering Simulation (NEESgrid)
- Grid-based Problem Solving Environment for Multiscale Flow Through Porous Media (FTPM)
- Grid-based Preliminary Aerospace Design & Optimization Environment (PrADO)
- Supercomputing in Plain English (SiPE)

Project	Who	OU Funding		Itanium Cluster Users				Status
		Source	Amt	Sr	Post	Grad	Under	
MEAD	KD,DW,HN	NSF/NCSA	\$150K	6	0	1	1	ongoing
LEAD	DW,KD	NSF-ITR	\$2.7M	6	0	2	2	preproposal sent
HEP	PS,JS	DOE	\$1.2M	8	3	6	5	ongoing
NEES	KM	NSF	\$870K	3	2	3	6	ongoing
FTPM	DP,HN	NSF-ITR	\$2.2M	5	2	7	8	proposal 2/12/03
PrADO	AS,HN	NSF, NASA	N/A	3	1	3	3	in preparation
SiPE	HN	NSF CRCD	\$400K	3	many	many	many	ongoing
TOTAL			\$7.5M	22	7	19	22	

All of the research projects above will use the Itanium cluster for essentially the same purpose: to design, develop, test, debug and deploy Grid-enabled software products and systems in anticipation of their release onto the Grid at large, and especially onto the TeraGrid. Some of the projects will develop Grid tools, some will develop Problem Solving Environments (PSEs), and some will use Grid computing to conduct computational science and engineering investigations.

Perhaps most importantly, the Itanium cluster will be used to teach both High Performance Computing and Grid computing to faculty, staff and students across a wide variety of fields, as an extension of OU's innovative "Supercomputing in Plain English" program, which targets not just experienced computing specialists but especially novice programmers with strong mathematics and science backgrounds. (Note that not all HPC and Grid computing education participants will necessarily use the Itanium cluster, but the Grid computing education initiative will be driven by experiences derived from its role in OU research.)

An important property of the proposed Itanium cluster will be its flexibility; that is, it will be maintained such that it can be reconfigured quickly and straightforwardly for unusual computing paradigms such as on-demand, real time and high throughput. This fact necessitates both a separate platform and a strong cooperative agreement among the PIs.

### **Meteorology: Modeling Environment for Atmospheric Discovery (MEAD)**

Research enabled by execution of simulation suites has a long and rich history; for example, the wind and thermodynamic conditions that lead to severe storms and produce large tornadoes have been explored this way. More recently, ensembles — simulation suites based on a variety of initial conditions, boundary conditions and models or model parameterizations — have become popular in forecasting. These simulations are then composited to provide probability forecasts.

With the advent of the TeraGrid, it is now feasible to conduct these simulation suites (with hundreds of simulations) and subsequent analysis/visualization at many sites. For example, atmospheric and ocean models running on different platforms can be linked effectively across the Grid. Further, simulation suites of coupled models are important for improving forecasting capabilities. Regrettably, conducting research that involves launching hundreds of simulations coupled with analysis, data mining, probability determination, and visualization (both individually and collectively) is daunting if each job must be submitted and catalogued individually, with resulting data analyzed, merged, and visualized in many painstaking steps.

Therefore, the goal of the MEAD expedition, led by the National Center for Supercomputing Applications (NCSA), is the development and adaptation of cyberinfrastructure that will enable simulation, data mining/machine learning, and visualization of hurricanes and storms utilizing the TeraGrid. The focus is on retrospective computation and analysis (not real-time prediction). Portal Grid and web infrastructure will enable launching of hundreds of individual Weather Research and Forecasting (WRF) [6, 7], Regional Ocean Modeling System (ROMS) [8], or WRF/ROMS simulations on the Grid, in either ensemble or parameter mode. Metadata and the resulting large volumes of data will then be made available through the MEAD Portal for further study and for educational purposes. The MEAD expedition will integrate model and Grid workflow management, data management, model coupling and analysis/mining of large ensemble data sets. It will fuse the ROMS and WRF codes, together with the Web-EH Portal [9] developed by OU's Center for Analysis and Prediction of Storms (CAPS), HDF5 [10] for data management, the Earth Science Markup Language (ESML) [11, 12, 13, 14, 15] for metadata coordination, and machine learning tools for intelligent data mining. Key elements of this work include workflow infrastructure, data management and ensemble services.

Workflow infrastructure supports coordination of job submission, scheduling, and collation; it will be coordinated with the Science Portals expedition [16, 17] and will build on the CAPS Web-EH toolkit. The existing Web-EH environment currently supports account and user login management, experiment design and input parameter configuration, and file transfers between the client, web and compute servers. Via an extended Web-EH interface, users will be able to specify model parameters for suites of ROMS/WRF simu-

lations. Because MEAD will generate a large number of simultaneous jobs when it is executing, an adaptive mapper will be used to assign these jobs to available Grid resources. This mapping will be based on technologies developed at Rice University that rely on performance models and system monitoring to ensure the distribution and scheduling of processes. This work will leverage the collaborations with the Performance Engineering expedition [18] to optimize the NSSL Collaborative Model for Multiscale Atmospheric Simulation (NCOMMAS) code [19]. In addition, XML [20] descriptions of model, analysis, and visualization parameters will be developed for entry into the data management system. Because such descriptions will be explicitly defined, they will be key to the integration and connection of the ensemble services. These XML descriptions will provide a mechanism for easily accessing data between MEAD components. By using these tools, the MEAD system will support the coordination of multiple jobs, both in terms of Grid exploitation and in terms of data movement between the jobs.

Supporting the latter requires a robust data management system. MEAD data management will leverage the Data Quest [21] and Scientific Portals expeditions and will require coupling of scientific and derived metadata with appropriate Data Grid [22, 23] metadata for seeking and acquiring data. Grave's ESML, based on XML, will be used for the interchange technology to handle the multiple data formats in MEAD. The ESML interchange technology effort will focus on coupling ESML tools with the Data Grid toolkit, including the GridFTP [24, 25, 26] protocol extensions. The data structural and semantic information described by ESML will complement the data content, location, and other information to be managed by the Data Grid toolkit. This will not only provide the MEAD system with access to heterogeneous data sources, but will also address the distributed nature inherent in linking multiple models and multiple tiled data sets. HDF5 will become a standard format for reading and writing WRF and ROMS data, which will improve portability and facilitate data and software sharing. Currently, WRF and ROMS access data in netCDF [27], so the addition of HDF5 access will provide methods for running simulations on large parallel systems.

The MEAD Ensemble Services will focus on the two community models (ROMS and WRF) as well as on a simpler finite difference model for exploration of ideas that could be incorporated into ROMS and WRF. ROMS has already been successfully coupled to mesoscale atmospheric models and can be run on either serial or parallel computer architectures, using coarse-grained parallelization that partitions the computational grid into tiles. These models will be coupled using HDF5 and components from the Model Coupling Toolkit [28]. When the ROMS and WRF are successfully coupled into an integrated modeling system, MEAD will be able to provide tools for analysis and learning of ROMS/WRF, as well as of ROMS and of WRF individually. These tools include the use of multiple data mining and machine learning approaches, including Welge's (NCSA) D2K [29] and Grave's ADaM mining systems [30, 31], which offer hundreds of mining algorithms, as well as the machine learning approaches used by the Grid-enabled parallel problems server (PPS).

PIs and Group Leaders: D. Weber, K. Droegemeier, (OU); R. Wilhelmson, J. Alameda (Univ. of Illinois at Urbana-Champaign); S. Graves (Univ. of Alabama at Huntsville); K. Kennedy, R. Fowler (Rice Univ.); D. Haidvogel (Rutgers Univ.); P. Husbands (Lawrence Berkeley National Laboratory); C. Isbell (Georgia Institute of Technology); P. Woodward (Univ. of Minnesota); B. York (Portland State Univ.)

Funding source, period and amount: NSF/NCSA; 10/01/2002 – 09/30/2003; \$149,999 (OU)

Status: ongoing

OU users: 6 senior personnel; 1 graduate student; 1 undergraduate student

### **Meteorology: Linked Environments for Atmospheric Discovery (LEAD)**

Every year, across the United States, violent, highly localized storms cause hundreds of deaths, disrupt transportation and commerce, and negatively impact the economy by \$300M a week. Although the economic and societal benefits associated with accurately detecting and predicting such mesoscale phenomena are enormous, the research and educational activities needed to achieve them are stifled by rigid informa-

tion technology frameworks that cannot accommodate the real time, on-demand, and dynamically adaptive nature of weather research; its disparate, high volume data sets and streams; and the tremendous computational demands of numerical models and data assimilation systems. Indeed, today's atmospheric research environments and related services are neither Grid-enabled nor web-based, both of which are essential for cutting edge weather analysis/prediction research and education.

In response to this pressing need for a national IT cyberinfrastructure in the atmospheric sciences, particularly one that can be interfaced with those of other relevant disciplines such as biology, ecology, the solid Earth sciences, and physics, we will address fundamental IT research challenges to create an integrated, scalable framework – known as Linked Environments for Atmospheric Discovery (LEAD) [32] – for use in accessing, preparing, assimilating, predicting, managing, mining/analyzing, and displaying a broad array of meteorological and related data, independent of format and physical location. A transforming element of LEAD is a dynamic infrastructure that will allow the operation of analysis tools, forecast models, and data repositories, not in fixed configurations or as static recipients of observational data, as is now the case, but rather as dynamically adaptive, on-demand systems that (a) change configuration rapidly and automatically in response to the weather; (b) are continually steered by data; (c) respond to decision driven inputs from the user; and (d) have the potential to drive remote observing systems. As a virtual extension of any user's desktop, LEAD will be available to all institutions regardless of size, and will provide mechanisms for individual users to create their own virtual servers for posting data and simulation results for use by others.

To meet these bold challenges, LEAD will create a series of interconnected, heterogeneous virtual Grid environments — pertaining to data, tools, connectivity, and services — that are linked at several levels to enable data transport, service chaining, interoperability, and distributed computation. These web-based environments will provide a complete framework within which users can identify, obtain, and work with observational and user-generated data in a distributed setting, and where the problem being addressed, the relevant data, and the computational resources can change with time and can be dependent upon or can control one another. In this model, the Grid workflow infrastructure will autonomously compute scheduling constraints, dynamically acquire resources, recover from errors and adapt to changing plans. A set of Integrated Grid and Web Services Test Beds will be developed for simulating and designing this dynamic Grid infrastructure. With them we will iteratively refine our thinking, leveraging extant software for data mining and management, Grid services, and monitoring and scheduling, develop new techniques for nimble adaptation, and deploy robust software for community use.

The University Corporation for Atmospheric Research (UCAR) Unidata Program [33], which is a LEAD partner reaching tens of thousands of university students, educators, operational practitioners, and researchers each year at more than 150 participating organizations, will serve as the focal point for deploying LEAD technologies to the atmospheric and environmental sciences communities. Specific emphasis will be given to research and education, with integration into formal operations following established paths. Education Test Beds will help shape LEAD research into applications that are congruent with pedagogic requirements, national standards, and evaluation metrics. By leveraging the many existing linkages between LEAD investigators and Minority Serving Institutions, traditionally underrepresented groups will be included in all aspects of LEAD research and education.

PIs: K. Droegemeier, M. Xue, D. Weber, K. Brewster (OU); S. Graves, R. Ramachandran, J. Rushing (Univ. of Alabama at Huntsville); R. Clark, S. Yalda (Millersville Univ.); R. Wilhelmson, D. Reed, M. Ramamurthy, P. Bajcsy (Univ. of Illinois at Urbana-Champaign); B. Domenico, A. Wilson, D. Murray (UCAR); D. Gannon, B. Plale (Indiana Univ.); E. Joseph, V. Morris (Howard Univ.)

Funding source, period and amount: NSF ITR; 09/01/2003 – 08/31/2008;

\$2,713,631 (OU request), \$14,946,098 (total request)

Status: preproposal submitted 11/18/2002

OU users: 6 senior personnel; 2 postdoctoral researchers; 2 graduate students

## High Energy Physics: Development of Data Grid Tools

Experimental particle physics faces enormous new challenges in the coming decade in the area of collaborative research. Experiments like ATLAS [34] and CMS [35] at the Large Hadron Collider (LHC) [36] at CERN will involve many thousands of physicists geographically dispersed all over the world. The traditional methods of conducting collaborative research — taking shifts at experimental facilities, commuting between national laboratories and universities, frequent workshops and meetings, and video or phone conferencing — are all inadequate for this challenge. We propose to develop a set of collaborative tools to enable groups of physicists to have effective wide area collaboration in particle physics.

General purpose experiments like ATLAS and CMS provide a stream of data that can be studied and tested against a variety of hypotheses. Traditionally, mining of this data (e.g., calibration, transformations, analysis) is organized through small groups of physicists, each working on a specific task or topic. These small groups carry out investigations over a specified period of time, and then report back to the full collaboration. If the full collaboration approves the results, it is presented to the wider community. In this working model, the majority of the work is done by small groups, from institutions all over the world. Traditionally, these groups have worked at the experimental laboratory, or have had to meet frequently and to participate in audio or video conferencing to carry out their work. We propose to develop a set of integrated collaborative tools to enable these collaborations remotely through audio, video and work environment sharing.

The OU/Langston High Energy Physics group, in collaboration with the University of Texas at Arlington (UTA), is involved in ongoing development of Data Grid [22, 23] tools for dynamic workspaces for scientific analysis communities. These Data Grid tools will allow small groups to create dynamic workspaces for their projects. We propose to create complementary collaborative tools that will allow people to share these dynamic workspaces interactively and in real time, empowering communities of physicists to effectively contribute to the LHC physics program while based at their remote institutions. This is especially urgent for the more than 600 U.S. physicists in ATLAS and CMS, since the LHC is located across the Atlantic.

The tools to be developed will create virtual collaborations in dynamic workspaces, where people will interact with Data Grid services in real time, while exchanging information via audio, video and whiteboards. Data will be manipulated and shared in real time through web services. For example, virtual control rooms will be created to manage large scale data production from remote locations. Collaborators will simultaneously access, manipulate and discuss graphical representations of selected physics data.

The groups at UTA and OU have played a leading role in developing tools for large scale Monte Carlo data generation using Data Grid middleware like Globus [37, 38] and Condor [39, 40]. We have developed Grid Application Toolkit (GRAT) [41] software, which is used for ATLAS data challenge production in the U.S. Thousands of CPU days are being made available for specific studies over short periods of 1-2 weeks using a distributed Data Grid involving 3 national laboratories and 7 universities in the U.S. Our first goal is to develop specialized “virtual control room environments” to manage such large scale productions. Remote collaborators will be able to view the same statistical, monitoring and control information from various sites while maintaining audio-visual communication. GRAT provides the tools to run production on the Data Grid from anywhere to everywhere. The collaborative tools proposed here will allow collaborations to share and manage the task of data production.

The OU/Langston group has developed web-based monitoring tools for D0 [42, 43] online data taking and triggering. For the D-Zero Online Systems group, Langston has developed several packages that provide remote access to online data monitoring. These four packages form a suite of tools that allow display of histograms generated by and controlling remote processes, enabling monitoring of the data as it comes in from the detector. These tools are written in C++ and built from the ROOT Object-Oriented Analysis Framework [44], taking advantage of ROOT’s graphical widget, network communications, threading, and object I/O classes. The tools implement a networked client-server paradigm, allowing geographical independence of monitoring activities, and consist of a multi-threaded Monitoring Server (MS), which is a component of the

monitoring programs; a multi-threaded Graphical User Interface (GUI) client that can control the remote monitoring processes or request information from them, including histograms; a single threaded Common Gateway Interface (CGI) client which acts as a proxy for the user's web browser when requesting the display of information from the MS; and a multi-threaded Server Registry (SR) that acts as a dynamic database of MS's to provide the clients knowledge about which monitoring tasks are currently running.

These tools can be adapted for use on the Data Grid using similar techniques to monitor and control data production. They are platform independent to the extent that ROOT is, except for the requirement that the CGI client must run on a host with an operating system that provides an X virtual frame buffer driver (e.g., Linux). The web interface to the CGI client is platform independent. Monitoring and control provided by these tools can be done from any location.

The tools described here will have enormous impact on the teaching mission at universities. Students will participate in real scientific challenges through these remote collaborations. Institutions with large minority enrollments, such as UTA, OU and Langston (a historically African-American university), will enable participation by traditionally under-represented groups in frontier physics experiments.

PIs: P. Skubic, M. Strauss, P. Gutierrez (OU); J. Snow (Langston); K. De, I. Ahmad (UTA); approximately 2000 collaborators worldwide

Funding source, period and amount: DOE; 04/01/2002 – 03/31/2004; \$1,190,000 (OU)

Status: ongoing; renewal submitted

OU users: 6 senior personnel, 3 postdoctoral associates, 4 graduate students, 3 undergraduate students

Langston users: 1 senior personnel, 2 undergraduate students

UTA users: 1 senior personnel, 2 graduate students

### **Earthquake Engineering: George E. Brown, Jr. Network for Earthquake Engineering Simulation (NEES) Computational Server**

This project will support the computational needs of earthquake engineers performing computational simulations under the aegis of the George E. Brown Jr. Network for Earthquake Engineering Simulation (NEES). NEES is a collection of experimental facilities for physical modeling of infrastructure, connected by a Grid-based fabric that forms the NEESgrid collaboratory, which links earthquake researchers across the United States with leading edge computing resources and research equipment, allowing geographically distributed collaborative teams to plan, perform, and publish their experiments.

The goal of this HPC project is to elevate computational simulation to serve as a peer to experimental simulation within the NEES project. Although the NEES systems integration architecture is Grid-based, there is no formal role for Grid computational simulation within the NEES project (though there is considerable demand for computational simulation methods within the NEES stakeholder community). One of the most important gaps identified by NSF within the NEES project is for the development of integrated scalable computational tools of utility to earthquake engineers. This proposal will have the specific impact of creating the first computational simulation site within the NEES project, and due to OSCER's unique charter, the success of this first computational site will readily be duplicated to help create a larger network of similar HPC resources deployed in support of the NEES project.

The scale of the proposed Itanium cluster hits the sweet spot of earthquake engineering computation, as it stresses large memory per node, a fast interconnect, and native 64-bit floating-point computation. These characteristics are fundamentally important to solving the general class of implicit unstructured (i.e., finite element) problems that commonly arise in earthquake engineering experiments, where a wide range of temporal scales must be modeled (e.g., the broad range of frequencies found in the earthquake response of critical infrastructure such as a long-span suspension bridge), and where the time scale of the analysis is so long that explicit temporal integration techniques are hamstrung by Courant conditions that limit effective modeling of the long solution duration. In short, most earthquake engineering problems are highly ill-

conditioned, with solution histories that span substantial periods of time; this proposal provides exactly the sort of high precision distributed memory supercomputing resources that are needed for effectively solving earthquake engineering problems.

The software resources that will be deployed on this Itanium cluster include Grid-capable open source finite element tools such as OpenSEES [45], high performance tuned applications such as DYNA3D [46], NIKE 3D [47], and the Diablo multimechanics framework (the OU team is a member of the LLNL Methods Development Group's Academic Partners program, so that these high performance codes are available for such use), and new computational tools such as the Terascale framework, which has been specifically designed for use within NEES on high precision Linux clusters.

The broader impact of deploying such computational resources in support of the NEES project is that earthquake engineering is an excellent metaphor for homeland security, because each involves unpredictable large scale attacks on critical infrastructure such as bridges, dams, buildings, and pipelines. Using HPC resources to achieve better hardened infrastructure designs through application of computational engineering techniques has a broad impact that is not limited to earthquake engineering, as this goal is fundamentally important to our entire society.

PI: K. Mish (OU)

Funding source, period and amount: DOE; 04/01/2002 – 06/30/2004; \$870,000 (OU)

Status: ongoing

OU users: 3 senior personnel, 2 postdoctoral associates, 3 graduate students, 6 undergraduate students

### **Chemical Engineering: Development of a Grid-based Problem Solving Environment for Multiscale, Multiphase Flow Through Porous Media in Hydrocarbon Reservoir Simulation**

Multiscale behavior of multiphase flow through heterogeneous porous materials is a complex physical problem that is very important in many kinds of applications: it is critical for understanding and predicting the behavior of a wide variety of systems diverse in function and scale, including hydrocarbon reservoirs, aquifers, separation tower and reactor units with packed beds, filters, catalytic converters, and even biological systems such as bone tissue. The specific area that our project focuses on is flow through hydrocarbon reservoirs, an area in which there has been a recent thrust to incorporate more physics [48], in order to improve reservoir productivity. The need for an improved understanding of flow and transport in porous media has become more critical at a time when the international landscape is in a state of flux and the U.S. needs to gain full advantage of technological advances in order to optimize the management and utilization of national resources. Furthermore, the possibility of either an accident or sabotage that changes reservoir properties in a sudden and dramatic way, and the analysis of the associated risks, should be realistically considered as part of the management of a reservoir.

Emerging HPC improvements will facilitate the analysis of flow through porous media, as well as the optimization of complex systems, using sophisticated approaches that have not been achievable in the past. Currently available simulations focus on only a single physical scale and on the determination of optimal well locations based on the initial conditions in the subsurface; the challenge is to combine multiple simulations, each addressing a different scale, into an integrated, multiscale simulator that can apply dynamic optimization of resources. Our objective is to develop a novel, robust, Grid-enabled, web-based software system, called "HiMUST" (Hierarchical Multiscale Simulator Technology), that can address many of these issues in a practical, user-friendly manner. Of great scientific and practical significance is the development of a simulator that will account for all physical scales and that will have flexibility in choosing the numerical method in accordance with the type of input data available, in choosing the optimization algorithm and objective, and in choosing the fundamental thermodynamic behavior of the fluid phases (e.g., run different components of the simulation depending on whether the input data are pore size and porosity measurements or digital images of rocks, provide a suite of optimization techniques and a suite of thermodynamic models).

Emphasis will be placed on portability, efficiency, and the Grid-enabled nature of the simulator.

The present project will combine emerging research needs in Grid computing with new research directions in engineering applications. The final software product will be a Problem Solving Environment (PSE) for the multiscale simulation of multiphase flow through porous media, which will also be rendered applicable to other physical applications by incorporating plug-and-play capabilities into the PSE. Specific objectives of the proposed project are:

1. Develop a Grid-enabled, web-based Science Portal encompassing an integrated multiscale PSE consisting of fully optimized and parallelized software components that simulate the behavior of a complicated physical system involving coupled as well as multiscale physics. Specifically:
  - (a) Develop a Science Portal based on the existing XCAT Science Portal system (XCAT-SP) [49, 50], to provide a Grid-enabled interface to the PSE for remote execution of multiscale flow studies by scientists and engineers across the nation (and, potentially, around the world).
  - (b) Integrate the software components to allow ongoing bidirectional, asynchronous feedback between the multiple physical processes and physical scales as a simulation progresses, thereby improving simulation accuracy by allowing not only for small scale properties to govern large scale phenomena but also the reverse, and in a dynamic, adaptive manner.
  - (c) Incorporate modules that allow plug-and-play extensibility of the PSE, in order to ensure that the PSE can be valuable for research or development in applications other than hydrocarbon reservoir simulation.
2. Develop a research tool that can be used to study the interrelation of complex physical processes. Specifically:
  - (a) Using an interconnected hierarchy of simulations for different physical phenomena (already developed or currently in development), study multiphase flow through porous media, with a focus on the interrelations between the physical scales and the effects of these relations on macroscopic reservoir behavior. The physical scales that will be addressed are: the macroscopic 3D reservoir scale behavior; flow in the mesoscopic scale, which will utilize a Pore Network Model (PNM) with a Monte Carlo approach [51, 52, 53], which has already been developed; multiphase flow at the pore level scale, based on the Lattice Boltzmann Method (LBM) [54, 55], which has also already been developed.
  - (b) Study multiphase flow through porous media for both low velocity (Darcy) and high velocity (non-Darcy) flow, including phase change and condensation close to gas wells.
  - (c) Develop methods for the quantification of uncertainty in the estimation of the porous medium properties, and for the effects of small-scale physical interactions on the macroscopic medium properties (e.g., non-Darcy flow coefficient, stress dependent medium permeability) and on the productivity of a reservoir.
  - (d) Optimize the behavior of the reservoir by dynamically managing the placement of the production or injection wells, and develop risk analysis procedures to account for uncertainties.
3. Incorporate information technology (IT) issues into Engineering courses, and incorporate scientific and engineering applications into both Computer Science courses and workshops with HPC content.
4. Transfer the resulting technology both to academia and to industry.

In the energy sector, both major and small independent companies are going to benefit from our effort. Major companies can utilize the PSE for various applications and will have access to the source code to improve and adapt to their particular needs. In industry, the budget for an effort comparable to the one described here would be larger by an order of magnitude. At the same time, such research and development work is outside the capabilities of smaller independent companies. However, they welcome technological advances that can improve their business, and the one proposed here will be very valuable to them. Furthermore, to achieve the goals of this project, the team must adapt the Alliance Grid Science Portal to work



with the several simulation codes, the optimization algorithms, and the thermodynamic models in several scenarios. Initially, this adaptation will occur simply, but as the project team explores the nuances of three mechanisms of code linking — unidirectional, bidirectional synchronous, and bidirectional asynchronous — it is expected that substantial improvements will be achieved in the underlying infrastructure for multiscale computing embodied in the Science Portal. Specifically, the bidirectional, asynchronous scenario will pose a variety of new challenges for the Portal architects, and will enhance the resulting Portal infrastructure; the XCAT-SP team has foreseen an upcoming need for this capability. In fact, the computational strategy that this approach embodies is expected to be much sought after among researchers whose applications require ongoing evolution of material properties. Furthermore, the improvements made to the Science Portal for this project will be a crucial experience for the Science Portal team and will furnish them with a much clearer conception of the nature and means by which cutting edge scientific applications operate in Portal environments. In this way, this project represents an ideal testbed for the Alliance Grid Science Portal in general and for a variety of enhancements to XCAT-SP that are anticipated to become crucial in the near future.

PIs: D. Papavassiliou, H. Neeman, M. Zaman, T. Trafalis, L. Lee (OU); R. Alkire, R. Braatz, J. Alameda (Univ. of Illinois at Urbana-Champaign); J. McLaughlin (Clarkson Univ.)

Funding source, period and amount: NSF ITR; 09/01/2003 – 08/31/2008;

\$2,220,000 (OU request); \$4,000,000 (total request)

OU users: 5 senior personnel, 2 postdoctoral associates, 7 graduate students, 8 undergraduate students

Status: proposal to be submitted 02/12/2003

### **Aerospace Engineering: Grid-based Design Environment for Preliminary Aircraft Design and Optimization**

The aerospace flight vehicle conceptual design phase is, in contrast to the succeeding preliminary and detail design phases, the most important step in the product development sequence, because of its predefining function. However, the conceptual design phase is the least well understood part of the entire flight vehicle design process, owing to its high level of abstraction and associated risk, its multidisciplinary design complexity, its permanent shortage of available design information, and its chronic time pressure to find solutions. Currently, the important primary aerospace vehicle design decisions at the conceptual design level (e.g., overall configuration selection) are still made using extremely simple analyses and heuristics. A reason for this scenario is the difficulty in synthesizing the range of individual design disciplines in a timely manner for both classical and novel aerospace vehicle conceptual designs, in more than an ad hoc fashion.

To efficiently complement and advance contemporary aerospace vehicle conceptual design approaches, an advanced aerospace vehicle synthesis system called PrADO (Preliminary Aircraft Design and Optimization Program) [56, 57, 58, 59] from the University of Braunschweig, Germany, has been acquired by the OU School of Aerospace and Mechanical Engineering (AME) as a platform for flight vehicle design research. Work is underway at OU's Aerospace Vehicle Design (AVD) Laboratory to generalize this code to be universally applicable to contemporary and emerging civil and military aircraft and space launcher projects. The in-house derivative development at the AVD Lab is called Aerospace Vehicle Design Synthesis (AVDS).

A key issue still to be resolved is computational turnaround time. Even with a robust state of the art vehicle synthesis system at hand, efficient investigations of complex multidisciplinary aerospace design problems are still constrained by lengthy computational runtimes. The traditional computing bottlenecks are the estimation of the aerodynamic database and weight estimation of the vehicle under investigation. In this context, the proposed research aims to significantly reduce the computational cost or turnaround time of this highly involved multidisciplinary vehicle design software, by integrating AVDS in a parallel computing environment in collaboration with OSCER.

Accelerating the evaluation of complex and highly coupled aerospace vehicle design tasks is key to enabling a design team to shorten the product development cycle, or alternatively to increase the quality

of the investigation, since a larger number of design studies can be performed in a given timespan. The proposed research undertaking will be instrumental in minimizing the ever present design risks inherent in aerospace vehicle conceptual design studies, by accelerating design response time.

Another fundamental issue in this context is the availability of this technology to a broad community of researchers and designers. In addition, flexibility and extensibility are critical aspects to the expansion of the design system beyond aerospace to other design fields. Therefore, development of a Design Environment (as opposed to a Problem Solving Environment) involving a Grid-enabled, web-based interface for developing data flows between individual design components, leading to flexible parameter space searches following user defined pathways, will be crucial to the useability of this technology. Because this project is in the early conceptualization phase, we will be able to incorporate Grid computing considerations into even the earliest software designs, rather than retrofitting our codes for Grid computing after their structures have hardened.

This research undertaking will have a profound value for the engineering community. Clearly, the core computer data management system of the generic aerospace vehicle synthesis software will be applicable to many industries. Having exchanged the individual analysis modules specific to the particular product under investigation (e.g., aerospace, automotive, ship), the data management system will be common across a wide variety of applications, allowing fast tracking of robust design convergency tests, disciplinary analysis, parameter surveys, and multidisciplinary single-point and multipoint optimization studies.

The development of a truly generic product synthesis system with a central parallelized computer data management system is the next logical step towards true multidisciplinary optimization (MDO) and design capability. The novelty of this research lies in the approach to a design system focused on the conceptual design phase of a product, in contrast to the current efforts that concentrate on design tools for the detail design phases of a product. Direct application to national security includes, for example, design of unmanned reconnaissance aircraft (UAV) and robots for rescue purposes. The ability to reduce overall design risks by coupling conceptual design with flight testing has a direct impact on operational safety. The contribution of the system to economics is outstanding, because of its effects on time savings and overall risk reduction. Finally, the development and application of the parallelized product synthesis system in an educational environment such as OU exposes students to state of the art MDO tools and HPC technologies, producing much sought after graduates.

PIs: A. Striz, B. Chudoba, H. Neeman (OU)

Funding source, period and amount: NSF, NASA; 01/01/2004 – 12/31/2006; amount to be decided

Status: proposals now in development

OU users: 3 senior personnel; 1 postdoctoral researcher; 3 graduate students; 3 undergraduate students

### **OSCER's Education and Research Strategy**

The OU Supercomputing Center for Education and Research provides the University of Oklahoma community with High Performance Computing infrastructure for facilitating education and research. Founded in the fall of 2001 as a division of OU's Department of Information Technology, OSCER has been developed as a cooperative effort among faculty and staff in 19 academic departments from 5 Colleges (Arts & Sciences, Business, Engineering, Geosciences and Medicine), working closely with OU's Chief Information Officer and Vice President for Research, and with enthusiastic approval from appropriate Deans and administrators.

OSCER's approach to HPC education and research [60, 61], partially funded by a National Science Foundation Combined Research/Curriculum Development grant ("Integration of High Performance Computing in Nanotechnology," \$400,000, 08/15/2002 – 08/14/2005), is unique, with its twin emphases on:

1. teaching HPC concepts, methods and technologies to researchers with substantial mathematics and science background but limited computing experience, through a series of workshops titled "Supercomputing in Plain English," and

2. mentoring via HPC “rounds,” in which OSCER personnel make regular one-on-one visits with each individual research team, helping them apply the HPC concepts that they have learned to advance their specific investigations.

OSCER’s educational efforts focuses on the following goals:

- Science and engineering students — both graduate and undergraduate — should be educated in Computational Science & Engineering (CSE) techniques, thereby positioning them for careers in cutting edge science and technology.
- Investigators whose research can be improved by effective use of HPC should be educated about HPC issues, hardware architectures, software technologies and problem solving environments.
- CSE researchers, not only faculty and staff but especially students, should be educated in the use of existing and emerging HPC technologies.
- Researchers who employ HPC in their investigations should be educated in software design methodologies that allow maximal exploitation of HPC resources.
- Researchers — including graduate and undergraduate students — should become aware of, and incorporate into their research codes, techniques of sound software design, coding, debugging, testing, maintenance, porting and performance evaluation.

To achieve these goals, OSCER provides a variety of HPC-oriented educational services:

- OSCER conducts frequent workshops on HPC issues, open to the entire OU community, and targeted not only toward active HPC-based researchers, but especially toward investigators with little or no HPC expertise but whose investigations could be enhanced by suitable application of HPC methodologies. These workshops cover HPC issues such as current hardware architectures, parallel programming paradigms, scalar optimization, existing and emerging HPC software technologies, performance evaluation, and Grid computing. In some cases, these activities serve to introduce HPC to new users; other presentations focus on the needs of experienced HPC consumers.
- OSCER assists interested instructors in coordinating coursework that has a direct relationship to HPC (e.g., Prof. S. Lakshmivarahan’s *Scientific Computing* course sequence in OU’s School of Computer Science), both to provide reliable access to HPC resources for use in pedagogical activities, and to assist in developing materials that expose students to HPC issues and that educate them in making effective use of HPC resources.
- OSCER provides web-based programmer education materials, targeted to student researchers in science and engineering fields, in order to help them to develop sound habits of software design, coding, debugging, testing, maintenance, porting and performance evaluation. (Some of these materials have already been developed for PI Neeman’s *Programming for Non-majors* course, which is targeted primarily to science and engineering students.)

OSCER assists in computational research in several ways:

- OSCER works directly with a wide variety of OU research teams to apply HPC techniques and technologies to their research. To date, OSCER’s HPC “rounds” have contributed to the success of over a dozen research groups, in such diverse areas as aerospace engineering, astronomy, business management, chemical engineering, chemistry, civil engineering, computer science, mathematics, mechanical engineering, meteorology, physics, and zoology.
- OSCER attracts researchers from disparate disciplines who have common interests, and fosters synergistic multidisciplinary collaboration, which enhances research teams’ ability to obtain external funding, an effect of critical importance in an EPSCoR state such as Oklahoma. For example, OU’s Data Mining group, which recently received an NSF Information Technology Research award, was formed as an offshoot of OSCER activities.

- OSCER assists principal investigators in developing funding proposals for computational research projects, providing not only a set of resources to list in proposals but also expertise on HPC issues that are of interest to funding agencies. OSCER especially targets untenured faculty (including Co-PI Papavassiliou), in order to assist them in obtaining the external funding, publications and broad recognition that are crucial to faculty success.
- OSCER cooperates with OU academic departments in the recruitment of high caliber faculty and students, including Co-PI Mish as well as the new Director of OU's Mewbourne School of Petroleum & Geological Engineering, Dean Oliver.
- OSCER works with participating researchers to develop appropriate paradigms for effective apportionment of resources, balancing the needs of multiple disparate groups in a fair, unbiased manner. Unconventional usage paradigms, such as real time and on-demand computing, are accommodated.

### **Additional Research and Education Projects**

In the early deployment phase, it is not expected that the Itanium cluster's resources will be fully consumed by the Grid computing projects described above. We will therefore make time available on the IA-64 Linux platform, at lower priority, to other computational research and education projects. While these projects may not focus on Grid computing, their use of the Itanium cluster will allow additional OSCER partners — studying dozens of research topics in a score of disciplines in biological sciences, business, engineering, geosciences, medical sciences and physical sciences — to advance to the next generation of commodity computing hardware. However, all non-Grid users will be required to accept that these resources can be withdrawn at any time, as Grid computing needs arise.

### **C. RESEARCH INSTRUMENTATION AND NEEDS**

Under this project, using a combination of NSF funds and OU cost share funds, OSCER plans to acquire:

1. A Linux cluster consisting of:
  - (a) a substantial number of Itanium processors (at least 16 and probably 32 to 64);
  - (b) substantial memory (at least 2 GB per processor and probably 4 to 8 GB per processor);
  - (c) the Myrinet [62] high performance interconnect;
  - (d) at least half a terabyte of local hard disk;
  - (e) a high performance connection to Internet2 (Gigabit Ethernet).
2. Considerable external storage:
  - (a) approximately 1.5 TB of permanent FiberChannel-1 hard disk, to be maintained as part of OSCER's IBM FAStT500 [63] disk server, which has already been deployed;
  - (b) approximately 20 TB of permanent AIT-3 tape storage, to be maintained as part of OSCER's Qualstar TLS-412300 [64] tape library, which has already been deployed.
3. Software necessary and appropriate for the Itanium cluster's planned user community:
  - (a) compilers for Fortran (77/90/95/HPF) from multiple vendors (Intel, Portland Group, Numerical Algorithms Group, GNU);
  - (b) the TotalView parallel debugger [65] from Etnus.
4. Significant system administrator time for the installation, deployment and ongoing maintenance of these resources during the 3 year project period. Specifically, we will appoint OSCER Manager of Operations Brandon George (biographical sketch in Supplementary Documents) as system administrator of the Itanium cluster, at  $\frac{1}{3}$  FTE during the project period.

Quotes for all of the above hardware, and for the TotalView debugger, can be found among the Supplementary Documents. These quotes should be considered as guidelines, because of the rapid price and capability changes associated with computing systems.

Other computing systems are available both at OU and in the region, but none of them will provide the needed capabilities that the proposed Itanium cluster will enable. OU currently maintains two HPC platforms: (1) a Linux cluster from Aspen Systems, Inc. consisting of 264 Pentium4 Xeon processors (2.0 GHz), 264 GB of RDRAM, the Myrinet2000 high performance interconnect, and multiple terabytes of local disk, and (2) an IBM p690 (“Regatta”) symmetric multiprocessor consisting of 32 POWER4 processors (1.1 GHz), 32 GB RAM and 200 GB of hard disk. Both of these systems are inadequate for the purposes of this proposal, because neither includes Itanium processors, and because both are capability platforms intended for compute jobs, rather than for development. For example, the CAPS real time weather forecasting system currently consumes about a quarter of the Pentium4 cluster, 24 hours a day.

On the other hand, the Texas Advanced Computing Center at the University of Texas at Austin has a platform very similar to the proposed system: a Linux cluster of 40 Itanium processors, 80 GB RAM and 720 GB of disk. However, this system is also inadequate for the purposes outlined above, because as a capability platform it lacks the flexibility to meet the evolving needs of the above research projects, and also, as a remote site, TACC’s mission cannot focus on the needs of these projects.

#### **D. IMPACT OF INFRASTRUCTURE PROJECT**

Each of the six research projects described above will derive considerable value from having an on-site Itanium cluster, which all will use for design, development, debugging, testing and deployment of Grid-enabled codes. In particular:

- The MEAD project will integrate two community codes, WRF and ROMS, to provide the numerical weather prediction research community with a dynamic workspace for ensemble weather forecasting and analysis.
- The LEAD project will establish a national infrastructure for high resolution, on-demand, realtime weather forecasting of evolving meteorological events.
- The High Energy Physics project will develop Grid tools for dynamic collaboration within a sizable but coherent research community, and these tools will prove useful not only within the HEP community but for others as well.
- The NEESgrid project will establish the first earthquake engineering computational site, for development of Grid-enabled computational strategies.
- The Flow Through Porous Media project will create a national computational resource for studying multiscale, multiphase flow phenomena across a broad variety of research areas.
- The PrADO project will create the first high performance conceptual design system and establish the first Grid-based aerospace design environment.
- The Supercomputing in Plain English project will expand an ongoing education initiative to include, for the first time, Grid computing education for scientists and engineers with modest computing background but strong need for these kinds of resources.

The particulars of how these goals will be reached can be found in the “Research and Education Activities” section, above.

The broader impacts of this project will be realized in several ways. First, and most clearly, these resources will enhance OU researchers’ ability to perform cutting edge Grid-based computational research — not only the investigators on this proposal, but also other research teams that participate in OSCER activities.

Second, the participating projects will act as testbeds for examining and enhancing High Performance Computing capabilities, exploring not only Grid technologies such as middleware, but also the most appropriate ways to perform computational research on Intel’s IA-64 processor architecture.

Third, OSCER has established a unique HPC education program, partially funded under an existing NSF CRCD grant, combining the “Supercomputing in Plain English” workshop series with HPC “rounds,”

thereby establishing partnerships between excellent scientific and engineering investigators and HPC experts, and leading to substantial gains in learning, research capability and output, and a growing community of computational research.

Fourth, this project brings young investigators from underrepresented populations to the cutting edge of high end computing, by strengthening an ongoing partnership between OU's High Energy Physics research team and its counterpart at Langston University, a historically African-American university, and through a new partnership between OU and Howard University via the LEAD project. In addition, OSCER has begun discussions with potential collaborators to develop a project for delivery of HPC rounds to minority serving institutions over the Access Grid.

Fifth, the research results generated by both the participating projects and other OU science and engineering teams will be broadly disseminated through journals and conferences. Included among these activities is an annual regional conference, conducted by OSCER and partially funded by Oklahoma EPSCoR, that in its first year saw over 60 participants, including many students, not only from OU but from other institutions in the state and region. The second annual OU Supercomputing Symposium is already scheduled for September 2003, and a significant increase in attendance is expected, especially among students.

Finally, OSCER is leading an effort among HPC sites in southern Great Plains, especially among EPSCoR states, to establish mutually beneficial interrelationships, particularly in the area of Grid computing. This project's resources will be leveraged to improve Grid-, HPC- and CSE-based research in nearby EPSCoR states.

## **E. PROJECT AND MANAGEMENT PLAN**

The OU Supercomputing Center for Education & Research is ideally positioned to take on the planning, purchasing, installation, deployment, administration and management of the resources described in this proposal. Since OSCER's founding in 2001, OSCER personnel have become experienced in, and adept at, these tasks, and existing structures and mechanisms are fully appropriate for this project.

OSCER is regulated by a charter that lays out the responsibilities of high ranking personnel. Governance is by a Board of Advisors, consisting of at least 9 and as many as 12 OU faculty and staff, who are elected as representatives of the members of the OU HPC community, which currently consists of over 100 faculty and staff from 19 departments in OU's Colleges of Arts & Sciences, Business, Engineering, Geosciences and Medicine. Five of the nine current Board members (Droegemeier, Papavassiliou, Skubic, Striz, Weber) are Co-PIs or senior personnel on this proposal, but the OSCER charter specifically requires annual turnover of approximately a third of the Board membership, so this situation will evolve.

The OSCER Board meets regularly, and also holds frequent e-mail discussions, in order to conduct OSCER business, set policy, and distribute resources. The Board's decisions are established by simple majority vote, and all meetings must have a quorum present (i.e., at least half the members) in order to take decisions. In many cases, decisions that must be taken on short notice, prior to the next Board meeting, are voted on via e-mail. To date, only one Board motion has not been unanimously accepted, and the concerns raised in regard to that motion have since been addressed by leveraging equipment money from a funded proposal.

The Board's decisions are executed by the Director of OSCER (PI Neeman), who is also responsible for day to day administration of OSCER business and who serves as a non-voting member of the Board. The Director also lays the groundwork for OSCER Board decisions, by researching purchasing issues, developing policy drafts for Board consideration, and investigating approaches taken by OSCER's counterparts at other institutions (e.g., computing resource allocation policies at national supercomputing centers). The Director reports directly to OU's Chief Information Officer, as well as to the OSCER Board.

While the Director focuses on policy, vision and initiatives, technical issues are the purview of OSCER's Manager of Operations, Brandon George, who is responsible for installation and maintenance of OSCER resources, and who reports to the OSCER Board on issues relating to his duties. Mr. George has been

developing expertise in the administration of Grid computing resources through his work with ongoing projects such as HEP and MEAD.

Allocations on the Itanium cluster will be made on a project by project basis, with high priority given to Grid computing projects and considerably lower priority given to other projects. Each project will be allocated resources, particularly computing time and storage space, based not only on need but also on the degree to which Grid computing is the project focus. Because of both the specific nature of the hardware resources and the well-defined focus of this proposal, the Itanium cluster will not be treated as a capability platform, but rather as a development environment, with a focus on flexibility and rapid response to evolving needs and paradigms.

## **F. RESULTS FROM PRIOR NSF SUPPORT**

The OU experimental High Energy Physics group has received \$343,100 over the last four years in NSF funds for R & D and construction of the ATLAS pixel detector. These funds were used to design, fabricate, and test three versions of prototype flex-hybrid circuits which are used to route signals and power between chips on the pixel module. The detector has pixels 0.02 square mm in area and is the ATLAS subsystem closest to the beam. The first working flex-hybrid pixel module was assembled and tested at OU. Additional modules have been studied in the CERN test beam and the expected position resolution (14 microns) has been obtained with 98.4% efficiency for tracks normal to the plane of the module.

Three institutions (OU, Langston University, and University of New Mexico) have been awarded \$513,195 (including 30% cost sharing) from the NSF MRI program for infrastructure equipment to complete construction of the ATLAS pixel detector. These funds will allow flex-hybrid assembly, and flex, sensor, and module testing to be done at our institutions. This will be a major contribution to the completion of the ATLAS pixel detector. (“Infrastructure for Semiconductor Detector Design and Readout;” award no. PHY-0116179 (UNM); \$513,195; subaward no. 3-20551-7820 (OU subcontract from UNM); \$153,269 for OU from NSF; cost share \$65,686 from OU VP for Reeach; 11/01/2001 – 07/31/2003). OU funds were used to purchase equipment for use at OU in design, assembly, and testing of flex circuits for the ATLAS pixel detector.

## References

- [1] I. Foster and editors C. Kesselman. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann Publishers, 1999.
- [2] D. Reed. Grids, the TeraGrid, and Beyond. *Computer*, 36(1):62–68, January 2003.
- [3] C. Catlett. The TeraGrid: A Primer. Technical report, TeraGrid, September 2002.
- [4] M. Cornea, J. Harrison, and P. S. Tang. *Scientific Computing on Itanium AE-based Systems*. Intel Press, 2002.
- [5] OSCER: OU Supercomputing Center for Education & Research website. <http://www.oscer.ou.edu/>, January 2003.
- [6] Weather Research & Forecasting (WRF) Model website. <http://www.wrf-model.org/>, January 2003.
- [7] J. Michalakes, S. Chen, J. Dudhia, L. Hart, J. Klemp, J. Middlecoff, and W. Skamarock. Development of a Next-Generation Regional Weather Research and Forecast Model. In *Proceedings of the Ninth ECMWF Workshop on the Use of Parallel Processors in Meteorology*. ECMWF, November 2000.
- [8] A. Shchepetkin and J. McWilliams. The Regional Ocean Modeling System: A Split-Explicit, Free-Surface, Topography-Following-Coordinate Ocean Model. <http://www.atmos.ucla.edu/~alex/ROMS/roms.pdf>, September 2001.
- [9] Web-EH website. <http://webeh.caps.ou.edu/>, January 2003.
- [10] NCSA HDF website. <http://hdf.ncsa.uiuc.edu/>, August 2002.
- [11] Earth Science Markup Language (ESML) website. <http://esml.itsc.uah.edu/>, January 2003.
- [12] R. Ramachandran, M. Alshayeb, B. Beaumont, H. Conover, S. Graves, N. Hanish, X. Li, S. Movva, A. McDowell, and M. Smith. Earth Science Markup Language. In *Proceedings of the 17th Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology, 81st Annual Meeting*. American Meteorological Society, January 2001.
- [13] R. Ramachandran, M. Alshayeb, B. Beaumont, H. Conover, S. Graves, X. Li, S. Movva, A. McDowell, and M. Smith. Earth Science Markup Language: A Solution for Generic Access to Heterogeneous Data Sets. In *Proceedings of the NASA Earth Science Techology Conference*, August 2001.
- [14] R. Ramachandran, , H. Conover, S. Graves, and K. Keiser. Interchange Technology for Applications to Facilitate Generic Access to Heterogeneous Data Formats. In *Proceedings of IGARSS*, Toronto, Canada, June 2002.
- [15] R. Ramachandran, H. Conover, S. Graves, and S. Christopher. Earth Science Markup Language: A Solution to Earth Science Data Format Heterogeneity Problem. In *Proceedings of the 83rd Annual Meeting*. American Meteorological Society, February 2003. To appear.
- [16] Scientific Portals Expedition website. <http://www.ncsa.uiuc.edu/Expeditions/Portals/>, January 2003.



- [17] XCAT-SP website. <http://www.extreme.indiana.edu/an/>, January 2003.
- [18] Performance Engineering Expedition website. <http://www.ncsa.uiuc.edu/Expeditions/Performance/>, January 2003.
- [19] NSSL Collaborative Model for Multiscale Atmospheric Simulation (NCOMMAS) website. <http://www.nssl.noaa.gov/~wicker/commas.html>, January 2003.
- [20] Extensible Markup Language (XML) website. <http://www.w3.org/XML/>, January 2003.
- [21] Data Quest Expedition website. <http://www.ncsa.uiuc.edu/Expeditions/DataQuest/>, January 2003.
- [22] Data Grid website. <http://www.globus.org/datagrid/>, January 2003.
- [23] A. Chervenak, I. Foster, C. Kesselman, C. Salisbury, and S. Tuecke. The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets. *Journal of Network and Computer Applications*, 23:187–200, 2001.
- [24] GridFTP website. <http://www.globus.org/datagrid/gridftp.html>, January 2003.
- [25] W. Allcock, J. Bester, J. Bresnahan, A. Chervenak, L. Liming, S. Meder, and S. Tuecke. GridFTP Protocol Specification. Technical report, Global Grid Forum GridFTP Working Group Document, September 2002.
- [26] H. Stockinger, A. Samar, B. Allcock, I. Foster, K. Holtman, and B. Tierney. File and Object Replication in Data Grids. *Journal of Cluster Computing*, 5(3):305–314, 2002.
- [27] R. K. Rew, G. P. Davis, S. Emmerson, and H. Davies. *NetCDF User's Guide for C: An Interface for Self-Describing, Portable Data, Version 3*. University Corporation for Atmospheric Research, June 1997.
- [28] Model Coupling Toolkit (MCT) webpage. <http://www-unix.mcs.anl.gov/larson/mct/>, January 2003.
- [29] Data to Knowledge (D2K) webpage. <http://www.ncsa.uiuc.edu/Divisions/DMV/ALG/d2k/>, January 2003.
- [30] Algorithm Development and Mining System (ADaM) website. <http://datamining.itsc.uah.edu/adam/>, January 2003.
- [31] R. Ramachandran, H. Conover, S. Graves, and K. Keiser. Algorithm Development and Mining (ADaM) system for Earth Science Applications. In *Proceedings of the Second Conference on Artificial Intelligence, 80th Annual Meeting*. AMS, January 2000.
- [32] LEAD website. <http://lead.ou.edu/>, January 2003.
- [33] Unidata website. <http://my.unidata.ucar.edu/>, January 2003.
- [34] ATLAS: A Toroidal LHC ApparatuS website. <http://atlas.web.cern.ch/Atlas/>, January 2003.
- [35] CMS: Compact Muon Solenoid website. <http://cmsdoc.cern.ch/cms/outreach/html/>, January 2003.

- [36] Large Hadron Collider (LHC) website. <http://lhc-new-homepage.web.cern.ch/lhc-new-homepage/>, January 2003.
- [37] Globus website. <http://www.globus.org/>, January 2003.
- [38] I. Foster and C. Kesselman. Globus: A Metacomputing Infrastructure Toolkit. *International Journal of Supercomputer Applications*, 11(2):115–128, 1997.
- [39] The Condor Project webpage. <http://www.cs.wisc.edu/condor/>, January 2003.
- [40] D. Epema, M. Livny, R. van Dantzig, X. Evers, and J. Pruyne. A Worldwide Flock of Condors: Load Sharing among Workstation Clusters. *Journal on Future Generations of Computer Systems*, 12, 1996.
- [41] GRid Application Toolkit website. <http://heppc1.uta.edu/atlas/software/user-tools.html>, January 2003.
- [42] D0 Experiment website. <http://www-d0.fnal.gov/>, January 2003.
- [43] S. Abachi et al. The D0 detector. *Nucl. Instr. and Methods*, A338(185), 1994. FERMILAB-PUB-93/179-E.
- [44] ROOT Object-Oriented Data Analysis Framework website. <http://root.cern.ch/>, January 2003.
- [45] Open System for Earthquake Engineering (OpenSEES) website. <http://opensees.berkeley.edu/>, January 2003.
- [46] DYNA3D website. [http://www.llnl.gov/eng/mdg/Codes/DYNA3D/body\\_dyna3d.html](http://www.llnl.gov/eng/mdg/Codes/DYNA3D/body_dyna3d.html), January 2003.
- [47] NIKE3D website. [http://www.llnl.gov/eng/mdg/Codes/NIKE3D/body\\_nike3d.html](http://www.llnl.gov/eng/mdg/Codes/NIKE3D/body_nike3d.html), January 2003.
- [48] J.W. Watts. Reservoir Simulation: Past, Present and Future. In *Proceedings of the SPE Reservoir Simulation Symposium*, pages 333–341, Dallas, 1997.
- [49] S. Krishnan, R. Bramley, D. Gannon, M. Govindaraju, R. Indurkar, A. Slominski, B. Temko, J. Alameda, R. Alkire, T. Drews, and e. Webb. The XCAT Science Portal. In *Proceedings of Supercomputing 2001*, 2001. <http://www.sc2001.org/papers/pap.pap287.pdf>.
- [50] D. Gannon, R. Bramley, G. Fox, S. Smallen, A. Rossi, R. Ananthkrishnan, F. Bertrand, K. Chiu, M. Farrallee, M. Govindaraju, S. Krishnan, L. Ramakrishnan, Y. Simmhan, A. Slominski, Y. Ma, C. Olariu, and N. Rey-Cenvaz. Programming the Grid: Distributed Software Components, P2P and Grid Web Services for Scientific Applications. *Journal of Cluster Computing*, 2002. Submitted.
- [51] H.-W. Lao, H. Neeman, and D. Papavassiliou. Stochastic Estimation of Porous Medium Properties. In *Proceedings of the 3rd International Conference on Computational Technologies for Fluid/Thermal/Chemical Systems with Industrial Applications*, 2001.
- [52] H. Neeman, H.-W. Lao, D. Simpson, and D. Papavassiliou. Multiscale Characterization of Porous Media Properties for Hydrocarbon Reservoir Simulation. In *Proceedings of the SPIE Conference on Commercial Applications for High-Performance Computing*, 2001.
- [53] H.-W. Lao, H.J. Neeman, and D.V. Papavassiliou. A Pore Network Model for the Calculation of Non-Darcy Flow Coefficients in Fluid Flow Through Porous Media. *Chemical Engineering Communications*, 2002. Submitted.

- [54] D. Grunau, S. Chen, and K. Eggert. A Lattice Boltzmann Model for Multiphase Fluid Flows. *Phys. Fluids A*, 5(10):2557–2562, 1993.
- [55] S. Chen and G.D. Doolen. Lattice Boltzmann Model for Fluid Flows. *Annu. Rev. Fluid Mech.*, 30:329–364, 1998.
- [56] Preliminary Aircraft Design and Optimization Program (PrADO) website. <http://www.tu-bs.de/institute/ifl/wolfgang.html>, January 2003.
- [57] H. Kossira, W. Heinze, and A. Bardenhagen. *An Integrated System for the Preliminary Design of Advanced Hypersonic Aircraft (PrADO)*. Springer-Verlag, Berlin, 1992.
- [58] D. Strohmeyer and R. Seubert. Improvement of a Preliminary Design and Optimization Program for the Evaluation of Future Aircraft Projects. In *AIAA Paper 98-4828*. Institute of Design Aerodynamics (DLR Braunschweig), 1998.
- [59] C.M. Oesterheld, W. Heinze, and P. Horst. Preliminary Design of a Blended Wing Body Configuration using the Design Tool PrADO. In *Proceedings of the CEAS Conference on Multidisciplinary Aircraft Design and Optimisation*, 2001.
- [60] H. Neeman, J. Mullen, L. Lee, and G. Newman. Supercomputing in Plain English: Teaching High Performance Computing to Inexperienced Programmers. In *Proceedings of the 3rd International Conference on Linux Clusters: the HPC Revolution 2002*. Linux Clusters Institute, 2002.
- [61] OU Supercomputing Center for Education & Research education webpage. <http://www.oscer.ou.edu/education.html>, August 2002.
- [62] Myrinet index page. <http://www.myri.com/myrinet/>, January 2003.
- [63] FASSt500 website. <http://www.storage.ibm.com/hardsoft/products/fast500/>, January 2003.
- [64] Qualstar TLS-4000 Tape Library website. <http://www.qualstar.com/146073.htm>, January 2003.
- [65] Etnus TotalView website. <http://www.etnus.com/Products/TotalView/>, January 2003.