Towards a Robust, Agile, and Comprehensive Information Infrastructure for the Geosciences:

A Strategic Plan For High Performance Simulation

Submitted to the

National Science Foundation

National Center for Atmospheric Research

University Corporation for Atmospheric Research

May 19, 2000

Boulder, CO

PREFACE

Strategic Plan for High Performance Simulation

We are pleased to submit this strategic plan for high performance simulation: *Toward a Robust, Agile, and Comprehensive Information Infrastructure for the Geosciences:* **A Strategic Plan for High Performance Simulation**. The creation of this plan was recommended by the National Science Foundation (NSF) Code Assessment Panel that visited the National Center for Atmospheric Research (NCAR) last summer. By preparing this plan now and by adopting a vigorous process of continuous review, we are establishing strategic priorities and principles that will guide us, both in the near term and in the coming decade, to meet the exciting challenges facing NCAR and the community.

The plan envisions an end-to-end simulation environment within NCAR and the University Corporation for Atmospheric Research (UCAR) Office of Programs (UOP), driven by scientific and technical needs and closely coupled to the broad scientific research community. We consider the end-to-end simulation environment, described in this strategic plan, to be an essential component of a needed comprehensive information infrastructure for the geosciences. The principal challenge at hand, of course, is scientific in nature, requiring that large-scale, complex simulations be conducted efficiently, effectively, and rapidly to facilitate and augment scientific progress in understanding the earth system.

Attendant to this principal challenge are three enabling challenges. These are: (1) how to design, acquire, and operate the required computational and related facilities to support the scientific community; (2) how to identify and meet the computer science challenges inherent in developing efficient algorithms to run on a variety of new and evolving platforms; and (3) how to effect a substantial change in the ways people from many disciplines and different locations work together to drive scientific progress.

We emphasize that this plan is strategic in nature. Its implementation will involve continued careful planning within NCAR and UCAR, in collaboration with the communities we serve and the NSF. Broadly based and substantive collaborations with the computer science and software engineering communities are essential for success. We view the plan as a living document that will be reviewed regularly, and that may be augmented and revised from time to time. However, it is our opinion that the major themes presented herein are sufficiently fundamental and general to serve us well for the better part of the decade.

The implementation of this plan is one of several needed steps toward the evolution of a comprehensive information technology strategy for UCAR. A strong foundation for this approach was laid by the UCAR Information Technology Council (ITC), which was created in 1997 to develop broad information technology strategies for the whole UCAR organization. The ITC plan, available at, http://www.fin.ucar.edu/itc/ITCStrategicPlan.html, specifies a wide range of activities that will contribute to the overall simulation environment. The rapid evolution of information technology presents UCAR and the community with many new opportunities to accelerate the progress in geosciences and to greatly change the way the community carries out its scientific and educational work. We embrace these changes and opportunities and look forward to working with the community to achieve the maximum rate of progress.

Initial steps for implementation will include presentation and discussion of this plan with the Geosciences Directorate and the Computer and Information Sciences and Engineering Directorate at NSF. We are also planning an internal workshop to be followed by a workshop involving participation by the geoscience and the computer science communities. The goals of these discussions and workshops are to define initial projects that incorporate the major themes in the strategic plan and to determine how the computer science and software engineering communities can and should participate in the implementation of the plan.

We are convinced that the successful implementation of this plan is essential for future progress in the atmospheric and related sciences and we are firmly committed to achieving of the goals described herein. We are also confident that these goals will be realized.

Robert J. Serafin, Past Director, NCAR

Richard A Anthes, President, UCAR

Timothy L. Killeen, Director Designate, NCAR

Acknowledgements

We thank the many people who participated in the thought and discussions that have led to this strategic plan. Their efforts have resulted in a document based soundly on research needs, technologic possibilities, and near- and long-term vision.

We begin by expressing gratitude to several members of the community who took time from their busy teaching and research schedules to review the plan, meet in Boulder for an in-person review of the plan, and make contributions regarding substance and style as well. These are:

Professor Otis Brown, University of Miami
Dr. Frederica Darema, National Science Foundation
Dr. Leo Donner, Geophysical Fluid Dynamics Laboratory
Professor Matthew O'Keefe, University of Minnesota
Professor Daniel Reed, University of Illinois
Dr. Richard Rood, NASA-Goddard
Professor Robert Wilhelmson, University of Illinois

The work of the NSF-appointed Code Assessment Panel has been identified elsewhere in this plan as influential in its development. It is also appropriate that we acknowledge Dr. Clifford Jacobs, our program officer at the NSF; we are grateful for his continuing dedication to UCAR, NCAR, and UOP in general, and the suggestions he contributed to this strategic plan in particular.

Acknowledgement and thanks are also due to the planning and writing committee members, who managed and developed the plan, and who have delivered a document that is rich in scope and ideas and which will serve the atmospheric and related sciences community well. They are:

Dr. Steven Hammond, Scientific Computing Division, NCAR, committee chair
Dr. Frank Bryan, Climate and Global Dynamics Division, NCAR
Dr. Peter Fox, High Altitude Observatory, NCAR
Dr. James Hack, Climate and Global Dynamics Division, NCAR
Dr. Joseph Klemp, Mesoscale and Microscale Meteorology Division, NCAR
Professor Robert Wilhelmson, University of Illinois
Dr. David Williamson, Climate and Global Dynamics Division, NCAR

Ms. Harriet Barker, Atmospheric Technology Division, NCAR, and Vice President Emerita, UCAR, assisted the committee in writing and editing matters.

Many other individuals participated in a series of planning discussions as the vision and strategies evolved. These are too numerous to name, but they include individuals at all levels in UCAR, NCAR, and UOP, from among the following groups: NCAR Scientists Assembly, NCAR and UOP Directors, NCAR Board of Trustees, the Information Technology Council, and all UCAR staff.

Table of Contents

Executive Summary

Advances in the geosciences are critically dependent upon simulation of the earth system, including solar-terrestrial interactions and atmosphere, ocean, sea ice, land, biogeochemical, and socioeconomic processes. The need for these complex and large-scale simulations, in turn, dictates that the geoscience community must have access to, and be able to use effectively, large-scale computational facilities.

NCAR and UCAR have essential roles to play, both in scientific research and in the provision of major facilities. Scientifically, NCAR will lead coordinated efforts in the development of earth system component models, coupled models, and their application to the most intellectually exciting and challenging scientific problems. On the facilities side, NCAR will provide high-end computational hardware, software tools, data storage and data management systems, data visualization tools, and a user-friendly computational environment. However, rapid changes in scientific needs and in information technologies dictate that the computational environment at NCAR both change and remain flexible in order for NCAR to maintain its position of leadership and to serve the university community. The need for enhancements and flexibility leads logically to the enabling vision of this strategic plan:

> *NCAR, in partnership with the NSF and the scientific community, will provide the leadership, resources, and expertise to establish and maintain a state-of-the-art high performance computational and simulation environment for use in addressing the most important and challenging problems in the atmospheric and related sciences.*

All parts of the scientific community have important and mutually supportive roles to play. NSF and other federal sponsors help to set the broad research agenda and provide the fiscal resources to carry out the research. NCAR provides scientific and computational expertise, facilities, collaborative leadership, services, and support, and serves as a virtual town hall for community planning and discourse. And UCAR, working with its university members, advisory committees, and the Board of Trustees, plays the major oversight role to establish broad strategies and to ensure that appropriate objectives are set, high standards are met, and resources are adequate.

An end-to-end simulation environment is necessary to achieve the vision. We define this environment as the entire domain within which scientific simulations are conducted and their results analyzed and disseminated. Such an environment provides a set of globally available resources and facilities that are beyond the scope or capability of a single institution to maintain. It also provides a set of tools whose benefits span many generations of project development. In the articulation of this strategic plan, six themes have emerged and each is considered fundamental to success.

THEME ONE	NCAR reaffirms its commitment to provide internationally recognized high performance computational capabilities for the atmospheric and related science community.

THEME TWO	NCAR will adopt software engineering practices that promote high performance and the efficient development of large simulation models and software infrastructure. NCAR will adopt or develop algorithms suitable for the simulation of geophysical systems and which can be efficiently implemented on highly parallel computer architectures.

THEME THREE	NCAR will provide data services and tools to manage and condense large data sets, extract and diagnose information, and convey or represent these in a form suitable for scientific interpretation.

THEME FOUR	NCAR will structure its management of major software development projects to accommodate multidisciplinary team efforts, led by project managers who have authority commensurate with their responsibility.

THEME FIVE	NCAR will augment its efforts in computer science, computational science, and software engineering to ensure a proper balance of these professions in the end-to-end simulation environment.

THEME SIX	NCAR will investigate, develop, adapt, and implement the infrastructure needed to support and facilitate effective collaborations in geographically distributed environments.

References to selected documents are provided throughout the text; these serve as indicators of the context within which this plan has been developed, and emphasize its status as a part of the national effort to capitalize on advances in scientific computing and information technology.

1 Introduction and Vision

1.1 The Vision

NCAR, in partnership with the NSF and the scientific community, will provide the leadership, resources, and expertise to establish and maintain a state-of-the-art high performance computational and simulation environment for use in addressing the most important and challenging problems in the atmospheric and related sciences.

This simulation environment will serve as the enabling mechanism for the broad research community to achieve rapid progress in the atmospheric, computing, and related sciences. It will also provide for the education and training of future scientists.

Scientific simulation is a rapidly evolving, fundamental tool for geoscience and environmental research, and the future is full of promise for continued, impressive progress. This plan presents a strategic view of how a community of people who focus on the challenging issues of earth sciences, defined here to include Earth, the Sun, and their interactions, can enhance its rate of progress with an appropriately balanced end-to-end simulation environment.

Simulations in the future will incorporate the complex relationships between humans and the earth system. The variability in the earth system on temporal scales ranging from microseconds to millions of years will be interpreted in a seamless way by those in need of timely, properly designed information for use in tactical and strategic decision making. The work to achieve the simulations will be carried out by a global community of scientists and students and it will be responsive to the needs of citizens, governments, and sponsors. The atmospheric and related sciences community is in a position to contribute in fundamental ways to scientific understanding, economic progress, environmental stewardship, and improved quality of life. This view and hope is consistent with UCAR's mission, as developed a decade ago:

> The UCAR mission is:
>
> To support, enhance, and extend the capabilities of the university community, nationally and internationally; to understand the behavior of the atmosphere and related systems and the global environment; and to foster the transfer of knowledge and technology for the betterment of life on earth.[1]

The community of participants—geoscientists, computational and computer scientists, social scientists, applied mathematicians, biologists, astrophysicists, biogeochemists, and ecologists to name a few—will participate as partners in an enterprise to take greatest advantage of current scientific understanding and the most advanced tools available.

Terascale computing at NCAR will be fully brought to bear on the largest and most important scientific problems of the earth sciences, a collection of grand challenge problems. A robust simulation environment will be integrated into a broader, high-bandwidth information technology infrastructure, which includes tools for software development, large-scale data management, visualization, and analysis. The integration of this functionality will become the end-to-end simulation environment we see as essential for providing rapid access to massive simulation archives and complementary observational data.

1.2 Environmental Simulation: A Grand Challenge

For many years, weather and climate modeling and geophysical and astrophysical turbulence simulation have been recognized as grand challenge scientific problems. In the 1980s, global environmental change was identified as the highest priority for environmental research and, following years of planning, the U.S. Global Change Research Program (USGCRP) and the International Geosphere-Biosphere Program (IGBP) were initiated. At its inception in 1991, the High Performance Computing and Communication (HPCC) program was organized by the NSF and other agencies to address these and other grand challenges. Since then, many studies by the National Research Council (NRC) and the NSF have placed environmental simulation, including but not limited to climate and weather modeling, as among the highest national research priorities.

Several national planning documents reflect the consistent appreciation of the need for advanced computing and its priority for U.S. science. These include the "Branscomb Report" in 1992[2], an NSF blue-ribbon panel report in 1993[3], a report by the NRC in 1995[4], and the President's 1998 budget proposal[5].

Of particular relevance is the recent report by the President's Information Technology Advisory Committee (PITAC)[6]. In the section of this report dealing with transforming our understanding of the environment, the committee says:

> *"Information technology can help us to improve a variety of problems from water and air quality to controlling the effects of toxic material. For example, reliable climate models permit us to determine the rate and regional distribution of climate change to support accurate projections by sector and region. Sophisticated models accurately predict the response of ecosystems to changes in temperature, water availability, and atmospheric composition. Fully integrated models allow scientists and policy makers to consider information on climate trends, population trends, resource utilization, and the value of natural and economic resources when making decisions regarding technically feasible and cost-effective options to reduce environmental impacts or adapt to climate change.*
>
> *To better support national and international energy and environmental policy, the United States requires an unprecedented acceleration and extension of research into climate modeling in order to improve the accuracy of local and regional forecasting. Progress in this area depends on improvements in computational methods. This will require orders of magnitude increases in computing capability to deal with the immense size of these problems in both time and space. We also need other advanced information technologies such as improved numerical methods and algorithms; tools for data storage, management, analysis, and visualization; software development and testing techniques; and advanced networks for distributed computing."*

1.3 The Role of Large-Scale Simulation in Understanding and Predicting the Environment

The role of large-scale computer simulation in understanding and predicting Earth's environment cannot be overemphasized. The environment is so complex and the interactions among the system's biological, chemical, and physical components so nonlinear, that simulation using numerical models is the only tool currently available for quantitative studies of the environment as a whole. Although traditional observational, theoretical, and process studies have important roles to play, simulation models are the tools that allow quantitative studies of the connections and feedbacks among the system's processes and components; testing of realistic hypotheses; predictions into the future; and assessment of mitigation strategies. In addition to simulation models, data assimilation techniques that incorporate diverse remote and *in situ* observations have been shown to be powerful tools in the synthesis or fusing of observations to describe the earth system.

Understanding our environment and predicting its future state on all time scales is an enormous intellectual challenge that requires terascale computing capabilities. Such capabilities include processor speeds, storage, input/output (I/O) and network bandwidth, and are about three orders of magnitude greater than those found on conventional university departmental systems. Table 1 summarizes the current status and near-term goals of sample earth sciences research topics, and provides a rough quantification of the computer capabilities needed to attain research model simulation goals.

One example will serve to illustrate how these quantitative estimates have been made; this is in solar magnetism studies, where scientists calculate convective heat flow around sunspots. The present version of this simulation code sustains 75 megaflops/processor on a particular cache-based microprocessor. At this performance level, a single calculation with the required spatial resolution of order 512^3 points must be run for at least 50 sound crossing times, or approximately $3x10^6$ time steps, and takes 104 days on a 1024-processor system. To run that same calculation in about a day thus requires a total of 7.3 teraflops from the same number of processors. Scientific simulations have commensurate needs for memory, disk capacity, visualization, and network bandwidth. For instance, the solar magnetism studies mentioned here require approximately 1 to 5 terabytes of disk space for primary variables and, optionally, an additional 2.5 to12.5 terabytes for secondary variables.

It should be noted that many of the research areas represented in Table 1 need tens to hundreds of sustained teraflops of computing capability. Because typical sustained rates are five to ten percent of peak, the simulations to support the relatively near-term goals of this research require computer systems with a peak performance of 100 teraflops to 1 petaflops.

Research Topic	Current Status	Goals	Computing Capabilities Needed to Reach Goals	Importance
Geophysical turbulence	Dynamic range limited to approximately 2.5 orders of magnitude, three dimensions	Dynamic range of 3 to 4 orders of magnitude, with ensembles	10–100 teraflops	Fundamental understanding of turbulence
Earth's middle atmosphere	Global scale, 1° resolution, 1–2 day simulations	5–10 km resolution, nested, 1–30 day simulations	10–40 teraflops for data assimilation and simulation	Understanding of fundamental processes
Operational regional weather prediction	Continental scale 10-30 km resolution, 1–2 day predictions	1–2 km resolution, nesting within global model, 2–5 day predictions	10–20 teraflops	Extend reliability, detail and value of predictions from 1–2 days to 5 days
Solar magnetism research models	Small active region and loop scale	Multiscale, full disk, active regions, corona, 100–2000 km resolution, nested, hour-to-day runs	20–100 teraflops	Fundamental understanding of solar surface features
Severe weather events research models	Simulate severe events 1–2 km resolution, 1–2 hours warnings	Resolution, 100–200 meters for 6-hour predictions	10–20 teraflops	6–hour advance warning of severe weather
Cloud-system simulations	Simulate convection, 2–3 km resolution, 10^6 km^2 domains, week-long runs	Resolve cloud-systems on continent/ocean basin to global scales; ensembles to derive probabilistic estimates; higher resolution	100 teraflops	Determine role of cloud systems in present climate and in modulating climate change
Space weather	Models separated by domain, short timescale, no prediction	Multiscale, nested resolution within global model, 2–5 day predictions	10–50 teraflops	Predicting the near-Earth environment
Operational global weather prediction	Prediction to 5 days at resolutions of 50–150 km	Ensemble predictions to 10 days at resolutions of 5–15 km	20–40 teraflops	Improved detail and reliability, value of predictions, to 10 days
Seasonal-to-Interannual climate prediction (El Nino/La Nina)	ENSO predictions, coupled atmosphere (200 km resolution) and tropical ocean (50–100 km) models, 90–360 days in advance	ENSO, monsoon, seasonal/interannual predictions, coupled atmosphere (30–50 km) and ocean (10–50 km) ensembles, 90 days to several years in advance	20–40 teraflops	Double the advance warning time and value of significant climate/weather forces having global economic impacts
Decadal to multi-century climate and anthropogenic change	Global models of ocean, sea ice, atmosphere, land (200–300 km), research and IPCC assessment simulations	Global models of ocean, sea ice, atmosphere, land, interactive chemistry, biogeochemistry, hydrology, cloud microphysics (10–50 km)	100 teraflops	Provide information to meet policy needs. Fundamental understanding of how climate works and life-climate system interactions

Table 1: Examples of Terascale Computing Needs in the Geosciences.

1.4 Community Roles

1.4.1 The National Science Foundation

The NSF has played a fundamental enabling role for the US university research community since the U.S. Congress created it in 1950. NSF provides essential funding to peer-reviewed programs of all sizes and—of equal or greater significance—participates with the national research community in setting future directions and identifying the most important research issues.

The National Science Board (NSB) has observed that NSF is among the larger supporters of environmental research in the United States. The Board further notes that NSF's investments, made on behalf of the US citizens, drive advances in fundamental understanding, and that NSF has designated research toward understanding and predicting the environment to be among its highest priorities for the years ahead. The NSF Geosciences Strategic Plan[7] and NSB reports highlight the importance of understanding complex nonlinear interactions of the biological and physical components of the environment.

> The first recommendation in one NSB report[8] is that "Environmental research, education, and scientific assessment should be one of NSF's highest priorities." Additional recommendations advocate increased "research on the methods and models that support the scientific assessment process" and on an effective infrastructure that includes "environmental observatories complemented by high-speed communications links, powerful computers, well-constructed databases . . . and both traditional and virtual centers that pull together interdisciplinary teams."

NSF's role in the future is as important and compelling as it has been during its first half century of existence: to lead and support national research efforts on cutting-edge scientific challenges of importance to society.

1.4.2 NCAR, UOP, the Universities, and UCAR

As a national center, sponsored by the NSF and with important support from other agencies, NCAR has a responsibility and a mandate to lead in achieving the ambitious national research goals entailed in understanding and predicting the earth system. NCAR brings to bear its observational systems, its capabilities in information technology and high performance computing, and its multidisciplinary scientific intellect to the grand challenge problems in the atmospheric and related sciences. From its inception in 1960, NCAR has a sustained history of open and productive collaboration with the university community in all aspects of its activities—research; management; facility development; planning; and model design, development, and deployment.

There are many examples of these collaborations. Some begin as research projects within NCAR and eventually become widely used community resources involving many university participants. The Community Climate Model (CCM) is of this type. It has evolved from its initial state in 1980 as a relatively simple atmospheric model to become the centerpiece of the fully coupled atmosphere-ocean-ice-land Community Climate System Model (CCSM). Others, such as the Penn State/NCAR Mesoscale Model begin in the universities. This Mesoscale Model has evolved from the low-resolution hydrostatic regional forecast model with simple physics it was in the early 1970s (MM1), to today's sophisticated nonhydrostatic multi-grid data assimilation and forecast system (MM5) with more than 600 users worldwide. (A brief summary of the community use of MM5 and CCM/CCSM is given in Appendix C.) A third example is an ongoing project to develop a new Weather Research and Forecasting (WRF) model. In collaboration with several government agencies and university scientists, this project seeks to provide a next-generation weather prediction system for use in both research and operations.

The UCAR Office of Programs (UOP) and its phalanx of education and support services have vital contributions to make. Important advances in support of scientists and students around the world include: developments in Unidata such as netCDF; the curriculum and digital library work in the Digital Library for Earth System Education (DLESE) Program Center; the training, outreach, and connections with operational weather forecasters embodied in Cooperative Program for Operational Meteorology, Education, and Training (COMET); the revolutionary use of GPS technology to provide a rich, new source of observational data for atmospheric sciences that has emerged from University NAVSTAR Consortium (UNAVCO); the global support of research programs and data management provided through Joint Office for Science Support (JOSS): and the distributed collaborations made possible through the Visiting Scientist Program.

The huge reservoir of talent and experience, encompassing the varied perspectives of faculty and students in universities-and researchers and technical staff in other institutions-must be fully integrated into our plans. Universities and other research and teaching institutions have strengths and diversity that complement those at NCAR and UCAR and vice versa. Collaborations among universities and NCAR have led to creative and effective work in conducting field programs, creating new instrumentation, identifying research priorities, developing simulation models, educating students at all levels, and establishing strong international relationships. Taken together, these capabilities represent a formidable set of skills and resources with which to tackle the high priority issues in the environmental sciences and in fulfilling the vision of this plan.

Finally, UCAR management, working with its members and advisory committees, and the UCAR Board of Trustees, provide oversight and review to establish broad strategies and to ensure that appropriate objectives are set, high standards are met, and that resources are adequate.

1.5 This Plan

This plan focuses on enabling and enhancing the science that relies on numerical simulation models coupled with high performance computing capabilities. The goal is to ensure that the atmospheric and related sciences community has available to it the most advanced computing and application support technologies possible. This document lays out the strategy for a new and augmented infrastructure necessary for an integrated, end-to-end simulation environment that will enhance and accelerate research on the principal grand challenge, which is to understand and predict the earth system. Embodied within this principal grand challenge are three enabling challenges. These are: (1) how to design, acquire, and operate the required computational and related facilities to support the scientific community; (2) how to identify and meet the computer science challenges inherent in developing efficient algorithms to run on a variety of new and evolving platforms; and (3) how to effect a substantial change in the ways people from many disciplines and different locations work together to drive scientific progress.

The first enabling challenge includes providing the advanced hardware to support the large modeling efforts. These facilities include, but are not limited to, fast processors, large random access memories, data storage devices, and fast, high bandwidth networks that connect scientists across the country and around the world.

The second enabling challenge includes the development of appropriate software engineering practices and the development of algorithms suitable for simulating geophysical systems on highly parallel computer architectures. As has been recognized for some time, the intrinsic nonlinear, highly interactive nature of geophysical systems poses fundamental computational challenges that will require innovative new approaches.

The third challenge involves changing the culture of the way people, who are necessary to achieve, overall success work together. Scientists from different disciplines often work in very different ways, use different disciplinary languages, and approach scientific and technological problems from different directions. Cooperation and collaboration are words that are easily said, but much harder to achieve in practice. Geographical separation further inhibits and complicates collaboration. These barriers can be lowered through advances in collaboratory technology.

A strong foundation and model for developing this plan was laid with the formation and activities of the UCAR Information Technology Council (ITC), starting in 1997. The ITC charge is to develop broad information technology strategies for the whole UCAR organization and implement them. An executive summary of the ITC strategic plan[9] is presented in Appendix A. This summary includes eight specific recommendations, the first six of which are particularly germane to this high performance simulation strategic plan.

Considering best practices and viewing science and technology challenges as opportunities led to the conception of the functional environment for high performance simulation. In examining and assessing the current balance and deficiencies in this environment at NCAR, six overall themes emerged. Together, the functional areas and specific activities described in the six themes will create the end-to-end simulation environment we seek to address the three enabling challenges. These functions are discussed in Section 2 and the themes in Section 4 of this plan.

2 The End-To-End Simulation Environment

The end-to-end simulation environment is defined as the entire domain within which scientific simulations are conducted. Such an environment provides a set of globally available resources and facilities, whose benefits span and are derived from many generations of project development. Progress in science that relies on numerical modeling is directly correlated with advances in hardware, software, and with the quality and effectiveness of an integrated end-to-end simulation environment. Procuring computer hardware is necessary but alone is not sufficient for creating the end-to-end environment. The crucial elements that make up a complete end-to-end environment are described below.

- **Hardware**
  The hardware component is a balanced set of computational resources: state-of-the-art floating-point and fixed-point processors, large random access memories, high speed memory interconnects, I/O subsystems, storage devices, multimedia devices, and networks, each with sufficient capability to satisfy the scientific objectives of large modeling efforts. Critical success factors are: commitment to sustained capabilities for addressing fundamental science problems, continuous reassessment of balance and scalability within the computing infrastructure, and priority-based allocation of resources among users and projects.

- **Software**
  Scientific simulations often require very large and complex codes that typically develop and evolve over a period of a decade or more. By contrast, the lifetime of a computer is typically only two or three years. The scientists, software developers, and the resulting codes may thus span three or four machine generations. Investments in model codes must be protected by suitable strategies for adapting to the rapidly varying technology areas. Critical success factors for software practices are: assessment of layered software frameworks; incorporation, development, and maintenance of critical expertise in software engineering methods and tools; numerical methods; planning for portability across the underlying infrastructure; and accommodation of the evolution of science requirements within an existing model framework.

- **Validation**
  Environmental simulations, analysis, and visualization–and the generation and sharing of new knowledge from them–are increasing in complexity. Thus, the testing, validation, and execution of all components become critical, especially in a diverse and distributed computing environment. A component-based software design strategy with well-defined interfaces facilitates validation. It also enables the simulation code and its pre- and post-processing stages to evolve together over the life of a project. Critical success factors are: incorporation of design reviews for all stages of the project, selection of appropriate measures of suitability for assessing simulations and their components, and development of suitable documentation of the validation steps.

- **Data**
  Data pass through all stages of observing systems and high performance simulations, and efficient handling of data is of fundamental importance to the environment. An effective strategy is needed to define simple abstract data models that hide complexities such as storage formats, organization, etc. Critical success factors for this data management are: inclusion of data-services handling in all framework-design aspects of simulation models; analysis and visualization; development and use of metadata conventions; coordination of distributed data archives; and specification, deployment, and support of tools and facilities for data services.

- **Analysis and Visualization**
  Managing and effectively analyzing and interpreting the very large datasets produced by simulations is fundamentally challenging. New and innovative methods for analyzing data are required in order to digest and understand the scientific content in the output of large, multidimensional models. Critical success factors for analysis and visualization are: development of analysis and rendering engines that enable researchers to locate, manage, resample, reduce, analyze, and visualize ensembles of simulation datasets; and adoption of frameworks that make use of simulation components.

- **Project Management**
  The breadth of current and envisioned modeling efforts mandates that greater attention be paid to the way science projects that require substantial software development efforts are managed. Project management must serve the science goals, identify key roles for the project, allow for a wide spectrum of collaborations and responsibilities, and carry the best practices learned to future projects. Critical success factors for project management are: ensuring that the science and research objectives are met by establishing model requirements and code specifications; assembling multidisciplinary project teams; identifying and effectively using resources; and developing robust, validated codes matched to the scientific objectives.

- **Community and Teams**
  Multidisciplinary teams and well-identified roles for each team and its members are necessary for implementation of the end-to-end simulation environment. Teams necessarily will involve people from several institutions, and their products—large simulation models, data, etc.—will be used by a wide community of scientists and students. In this context, institutional commitment to and support of such models as community facilities is essential. Critical success factors are: assessment of community needs, commitment of appropriate resources, development and maintenance of appropriate documentation, outreach and training for major model projects, specification and development or adoption of scalable tools for recipients of educational services, and leverage of existing expertise related to these services.

3 The Computational Operating Environment

3.1 Interdisciplinary and Multi-Institutional Models

The earth system is highly interdisciplinary with many complex interrelationships. The complexity and diversity of this system is graphically illustrated in Figure 1. It shows a version of the "Bretherton Wiring Diagram," modified by Brasseur, to include an enlarged human dimension component and augmented interaction arrows to better represent current understanding of the interactions among components in the system.

The structure of major environmental sciences modeling efforts reflects the complexity in the earth system. They have become broad, interdisciplinary and multi-institutional endeavors. Current coupled climate models incorporate atmosphere, land, ocean, biosphere, and sea ice components. In the Community Climate System Model (CCSM) project, development of the atmospheric component was originally centered at NCAR. Current development activities are carried out collaboratively with scientists at the National Aeronautical and Space Administration (NASA) Data Assimilation Office, Oak Ridge National Laboratory, and Lawrence Livermore National Laboratory. Development of the ocean component is conducted jointly by NCAR and Los Alamos National Laboratory (LANL), development of the sea ice component is primarily centered at LANL and the University of Washington, and development of the land component is distributed among the University of Texas, the University of Arizona, and the Center for Oceans Land and Atmosphere (COLA) and NCAR. The marine and terrestrial biogeochemical component models (to be incorporated into the system in the near future) are being developed at NCAR, the University of Wisconsin, Woods Hole Oceanographic Institute, the University of California at Berkeley, LANL, and the Max Planck Institute for Biogeochemistry. The new Weather Research and Forecasting (WRF) model is being developed in partnership among many university scientists, NCAR, the National Oceanic and Atmospheric Administration, and the Department of Defense. The Comprehensive Space Environment Model (CSEM) couples models of the solar corona and solar wind with a model of the Earth's magnetosphere. It is funded under the NSF Knowledge and Distributed Intelligence initiative and is jointly under development by the University of Michigan and NCAR collaborators.

Researchers face formidable political, scientific, and technical challenges in coordinating the distributed collaborations that are now commonplace. Forming effective teams comprising individuals with differing interests, expertise, organizational affiliation, communication patterns, and promotion and reward systems adds considerable complexity to the management and conduct of projects. From a software engineering point of view, challenges include software interoperability, code sharing, software version control, validation and testing, and data set integration. The ambitious scientific goals of the modeling efforts mentioned here can only be achieved through adoption of effective project management and coordination and sound software engineering practices.

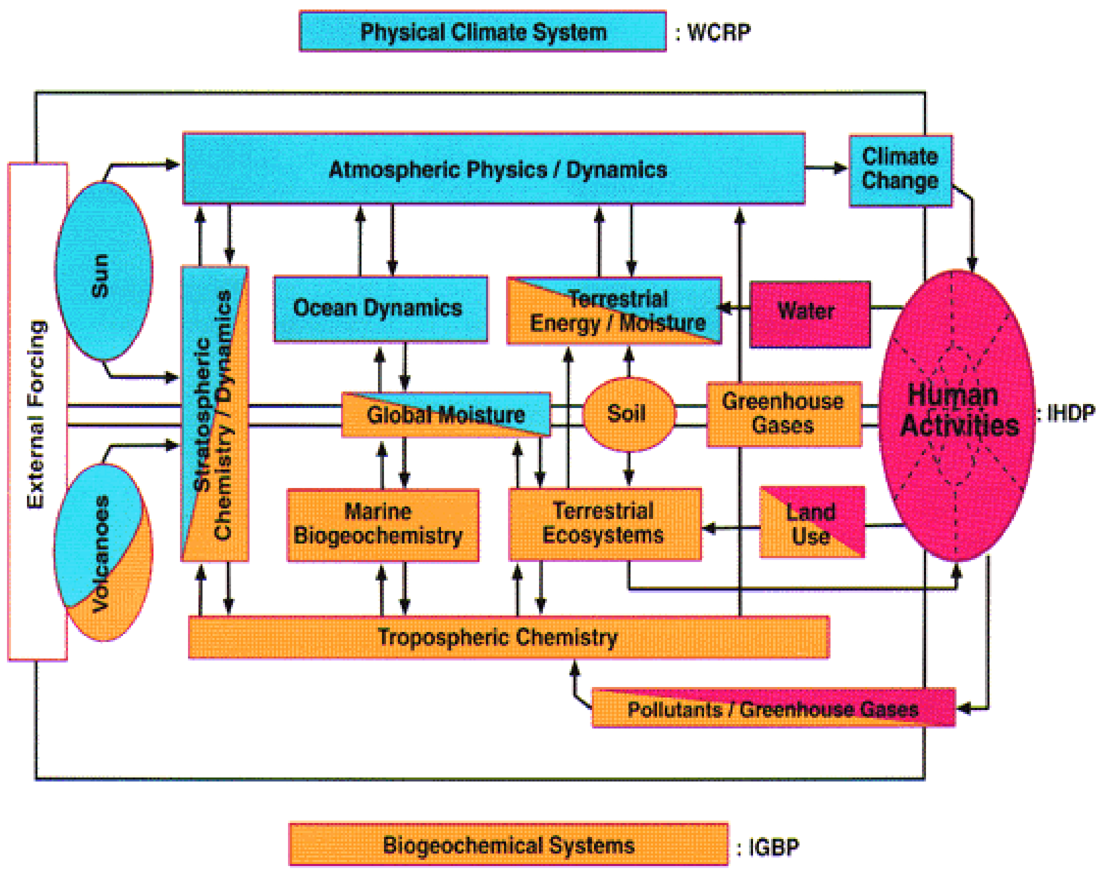

Figure 1: The Earth System. A modified "Bretherton Diagram" highlighting linkages between social (red), biogeochemical (orange), and physical climate systems (blue). Courtesy of Guy P. Brasseur, Director, Max Plank Institute for Meteorology, Hamburg, Germany.

3.2 NCAR and the National Computing Environment

Not only are models being developed through multi-institutional collaborations, they are also being run at a variety of facilities. For example, NCAR's scientists are currently using the Parallel Climate Model (PCM) for simulations at NCAR, as well as at four different Department of Energy (DoE) laboratories, on systems with three relatively disparate architectures. This trend is likely to continue. Therefore, models have to be engineered to achieve good performance, and also to be portable across a variety of architectures and across a variety of vendors' systems.

There are tradeoffs and costs associated with portable models. For example, the performance on any particular system may be somewhat lower for a portable model compared to one that is developed and tuned for only one system. Additional effort is often required in order to achieve portability. This additional effort will be offset by increased opportunities for access to computing cycles on a variety of machines at multiple facilities, and by achievement of a more robust code.

In the 60s and 70s, high performance computing and data access at NCAR required researchers to journey to Boulder. There were no networks or personal computers to carry out research. This has changed dramatically in the past two decades, and NCAR computing and data facilities are now accessible over commodity and research networks. In addition, data sets and modest-to-very-powerful computing facilities have become pervasive throughout the research community. Scientists and students are not constrained to use a particular center, computer system, or network. Collections of these resources, including those available at NCAR, are needed to carry out today's ambitious and interdisciplinary research agenda. Use of these collections of resources is enabled through development of "virtual machine rooms" where components are coupled through a network or "Grid" fabric. The Grid (as it has come to be called in the research and commercial sectors) can generally be defined as a heterogeneous complex of advanced computers, networks, storage and display devices, and instruments. The Grid fabric includes advanced networking services such as resource management, accounting, communication, security, and remote data management. Integrating NCAR's resources and utilities into and providing access to this Grid is a further challenge for the community in general and NCAR in particular. It is a critical part of the end-to-end simulation environment.

NCAR must consider community and research codes and associated services at all organizational and community levels, not just with respect to large climate and weather models. Both the simulation models and the end-to-end simulation environment should be designed to permit simulations to migrate directly to other larger-scale facilities, such as those provided in NSF's Partnership for Advanced Computational Infrastructure (PACI). In addition, we must develop relationships with these centers so that the end-to-end simulation environment at NCAR is compatible with them. Overall, models and services will span many platforms; others will fall into, or be targeted to, specific capability classes–e.g., desktops, workstations, or supercomputers. This diversity guides the development of the associated software architecture, which must be flexible, portable, and coordinated with other centers.

3.3 Volatile U.S. High Performance Computer Industry

The current U.S. High Performance Computing (HPC) environment increases the complexity of building advanced scientific models. The market for these systems is small and volatile, and no single vendor is dominant. Furthermore, the market is characterized by rapid reversals of fortune, and attendant fluctuations in the level of vendor commitment. The prevalent architecture—a parallel cluster of shared-memory multiprocessor nodes built with commodity parts—is perhaps the most complex computing system ever created. Multiprocessor systems typically have deep memory hierarchies and complex caching schemes. The most common programming models are either virtual shared memory or a hybrid of shared and distributed memory.

Achieving high performance on these architectures requires a careful combination of caching optimizations, shared memory directives, and message passing calls. The diversity of the architectural characteristics of present and future computing systems dictates a flexible, layered software design that will provide good performance across the spectrum of known architectures, and that can be adapted readily to future innovations in computing hardware.

Not long ago, HPC vendors offered well-integrated systems including debuggers, performance analysis tools, and queuing systems. However, it is becoming more and more common for vendors to rely on third parties to provide such software. While this has the potential benefit of providing a common user interface to tools and user environment across multiple systems, the level of integration for multiple software packages varies and the installation and maintenance increases the burden on system administrators. The trend toward more commodity computing systems has given rise to clustered systems, called "Beowulf" systems. Typically, these are homegrown parallel systems running the open source Linux operating system; they can offer comparable performance to vendor-delivered systems at a fraction of the cost.

3.4 National Focus on Advanced Computing and Infrastructure

The evolution of NCAR's simulation-based research as described in this strategic plan is aimed at improving the quality of the software development efforts at NCAR and elsewhere in the community. Another goal is to protect existing software investments and assets, and reduce the difficulty of model development through modular code design and code reuse. This plan is well aligned with new NSF initiatives in Information Technology Research and the draft NASA Cooperative Agreement Notice (CAN) from the HPCC program, which focuses on "Increasing Interoperability and Performance of Grand Challenge Applications in the Earth and Space Sciences."

The issues in high performance computing have become too diverse and complex for individual organizations to deal with on their own. Institutions with common interests and needs must join forces to leverage their expertise and resources. In achieving the desired computing environment of the future, UCAR must continue and expand its commitment to creating strategic alliances with universities, other centers, government labs, and industry. Strengthening and expanding the relationship between NCAR and the National Computational Science Alliance (NCSA) and the National Partnership for Advanced Computational Infrastructure (NPACI) is important to our success. NCSA and NPACI are supported by the Partnerships for Advanced Computational Infrastructure (PACI) at NSF and include university and industrial partners.

4 Strategic Plan Themes

Conducting simulation-based research on terascale computing platforms carries with it enormous challenges of scalability and adaptability. Discussions with scientists and programmers who currently are involved in such simulation projects contributed to our understanding of the scope of this challenge. These discussions led to the formation of a functional environment for high performance simulation. In examining and assessing the current balance and deficiencies in this environment at NCAR, six overall themes emerged. These six themes address the highest priorities and together with the functional areas create the end-to-end simulation environment. Many aspects of the themes have been embraced by various portions of large modeling groups at NCAR. However, they have not been uniformly adopted throughout UCAR; therefore this is now an important goal for the organization as a whole.

4.1 Theme One: Computing Resources

NCAR's mandate as a national center includes providing computing resources and addressing problems that require sustained high performance. Current high-end simulations require computers with sustained teraflop performance, and commensurate memory, I/O, and network bandwidth, and a rich software development environment. NCAR must seek to provide a robust high performance computing capability to meet scientific simulation needs, independent of vendor, machine architecture, or programming paradigm. NCAR's strategy must serve a broad range of community needs, with a special emphasis on grand challenge, leading-edge science. Some components of the end-to-end environment, such as the Mass Storage System, will need to service a continuum of small-to-large projects, and systems must be designed and function with this continuum in mind. Much of the analysis, development, and validation work associated with software projects requiring high-end capabilities will be conducted on desktop-to-intermediate-performance systems, and these systems must be well integrated into the end-to-end environment.

4.1.1 Computing

THE PLAN: NCAR reaffirms its commitment to provide an internationally recognized, high performance computational capability for the atmospheric and related science community.

Although NCAR still provides a formidable computing capability, the facilities and components of the supporting infrastructure have fallen behind peer centers. Thus, the rate of scientific progress for scientists and students using NCAR facilities, where simulation is the primary tool, has been unnecessarily limited by inadequate sustained computational capability. To reinvigorate these research activities, NCAR and the community must immediately seek an order-of-magnitude increase in sustained computational capability, followed by regular upgrades. The geosciences community has mature and well-established scientific research programs ready to use the maximum amount of computing power available.

THE PLAN: NCAR will increase its participation in Grid activities with particular attention to needs within the atmospheric and related sciences.

As mentioned previously, various groups within the research and commercial sectors are investigating the development of "virtual machine rooms" where components are coupled through a network or Grid fabric. They are also building and studying the services and "middleware" that are needed. Various federal efforts, including NSF's Partnerships in Advanced Computational Infrastructure, NASA's Information Power Grid, and DoE's Next Generation Internet program, are providing support for these efforts. An integrated Grid Architecture is being defined to enable scientists to carry out research in the 21st century. With respect to NCAR, there is a need to define and implement application architectures that sit on top of the Grid middleware, as well as to make NCAR-specific resources Grid-functional. This will enable researchers to use computing resources at multiple sites readily and effectively. Further, model data from distributed sites, together with data from observational programs, should be readily moveable through the Grid for use in data assimilation, ensemble and model-observational comparison, analysis, and visualization.

4.1.2 The NCAR Mass Storage System

THE PLAN: NCAR will invest increased resources for Mass Storage System facility development at the service layer.

Providing adequate computing capability is critical to the end-to-end simulation environment, but it is not sufficient. Reliable and capacious storage systems are required as well. For some time, NCAR's Mass Storage System (MSS) has provided a substantial reservoir for data holdings. At the time of this report, the MSS holds approximately 250 terabytes of data and more than 7 million files. Mass store capability must grow in capacity, accessibility and usability commensurate with the future growth in volume of data that will be generated by terascale simulations. Capacity and performance enhancements are an ongoing process. Historically, there is an approximately linear relationship between the amount of MSS data transferred and stored and the amount of computing capacity at NCAR. Technology will continue to provide faster storage devices and denser media that will need to be incorporated into the MSS. For example, relative to the current technology in the NCAR MSS, we expect an approximate fourfold increase in data transfer rates and an approximate sixfold increase in media density for tape storage devices in just a few years.

In addition to ensuring timely performance and capacity enhancements, augmenting MSS functionality and services will also be necessary for a balanced end-to-end simulation environment. As part of this, NCAR will provide Web-based tools to access and manage MSS data and metadata. In addition, NCAR must shift its focus from the data files to data services (see Section 4.3.1). Required services include an interface for abstract data types, including transparent handling of parallel I/O, metadata creation and handling procedures, and automated population of catalogs. A high capability, high performance shared file system is also required for effective provision of services in a distributed environment. Finally, advances in the community and industry will be tracked and evaluated for deployment at NCAR. For instance, the Grid community has pioneered the emergence of the Grid Security Infrastructure with a standard mechanism for accessing remote storage systems.

Another expansion of the MSS functionality will be the MSS File Services, which will provide data sharing in a heterogeneous environment. These file systems will enable drag-and-drop interfaces, as well as data sharing among a heterogeneous set of computing servers, data analysis servers, visualization servers, desktops, and Web clients. The MSS File Services will be deployed in a distributed fashion, each component working to provide a subset of the total data name space. Considerations for distributing MSS File Services include physical distribution in order to support the multiple UCAR sites, UCAR divisional File Services, and university site access.

4.1.3 Networking

For general-purpose computing, commodity networking is an essential and enabling technology. UCAR has provided robust and reliable networking to allow the organization to function and prosper in a rapidly evolving technological environment. In the context of the end-to-end simulation environment we envision here, particularly the high performance segment, networks must be able to support the research demands, offering not just theoretically high wire speeds, but being able to achieve close to maximum throughput in a sustained manner, much as is the goal for processors. They must also offer a high level of redundancy and fail-over, since the sheer volume of information to be transported may preclude re-transmission.

Networks allow research calculations on distributed supercomputers, data assimilation with analysis of terascale datasets, distributed visualization, and remote collaboration among scientists and students. Because of the range of individual and collective needs, these networks, consisting of at least primary and secondary backbone conduits, must include sufficient bandwidth, be designed for reliability, ensure data transfer integrity, and be scalable. Often, transfers of information require a burst-mode capability that far exceeds the mean transfer rate for which the network has been designed. This burst mode is critical for distributing datasets to local data service centers and visualizations to desktops, and for exchanging information streams with distributed archives.

> *The PLAN: NCAR will actively engage in state-of-the-art networking research projects to support the end-to-end simulation environment.*

Fundamental problems still exist in networking research, such as implementations of lower layers of the basic transport protocols for maximum efficiency. As an example, a research project in which NCAR is currently participating is known as WEB100. Its aim is to establish a complete host-software environment that will run common Web applications at 100% of the available bandwidth, regardless of the magnitude of a network's capability. In particular, applications would be able to automatically consume 100% of the available bandwidth in very-high-performance networks, something that has been problematic to date in today's host operating environments. The fundamental strategy will be to drive what has become commodity technology and software into the high performance category.

In the end-to-end simulation environment, many or all of the application components and tools will be network aware, and will be required to handle audio, video, encapsulated objects, etc. in addition to regular file and message transfers. Research into network tuning and network access to distributed resources via more intelligent routing and switching technology is required.

Further, in the development of Beowulf-class microprocessor clusters, networking technology, in the form of switching and multiplexing, can quickly become a limiting factor in the efficient utilization of these systems for high performance simulation applications. Research is required in an operational environment to identify optimal configurations as well as what fundamental limitations exist.

THE PLAN: NCAR will remain well connected to regional and national networks to disseminate scientific results effectively.

Fast local, metropolitan, and wide-area networking is also a critical component of the end-to-end simulation environment. The services offered within a community must span the usual geographical limitations and extend multimedia connectivity to allow collaborations to flourish. NCAR must have a presence in those research networks that support the multidisciplinary teams. The ability to meet these requirements and deliver services, especially at high performance, is often limited by network security restrictions imposed by exposure of research networks to commodity patrons; these needs and restrictions must be assessed by the community of participants. With the emergence of the Web browser as a common window to many successful commercial, database, scientific, and educational services, NCAR must consider researching and utilizing or jointly developing browser plug-ins, as well as providing remote desktops and other working environment tools over distributed networks.

4.2 Theme Two: Software Tools, Frameworks, and Algorithms

NCAR places a premium on developing high performance, portable codes that are shared and reused, and that insulate the modeling and scientific efforts from hardware changes. One way to achieve this is through the use of object-oriented and object-based software design, components, and frameworks. These can provide an effective modeling infrastructure that streamlines the development of complex software. The Glossary and Appendix B contain definitions for each, but in general, components allow code reuse, while frameworks promote design reuse. Design reuse has advantages over code reuse since it can be used in more contexts and can be applied earlier in the development process. These reuse techniques have inherent short-term costs; they require analysis to extract the commonality from the individual applications and require coordination to design common solutions. In addition, there are costs to create and maintain documentation and software repositories. However, well-designed software frameworks that meet performance objectives will pay large dividends to major modeling efforts.

4.2.1 Software Engineering Practices

THE PLAN: NCAR will adopt software engineering practices that promote high performance and the efficient development of large simulation models and software infrastructure.

Simulation has emerged as a fundamental method for research; accordingly, we must ensure that scientific software meets the same standards of practice as do design and planning of laboratory instrumentation and field or lab experiments. The commercial sector, where development of complex software systems is a driving force, employs software engineering practices that are not routinely used in scientific research environments. NCAR will evaluate approaches to software development to determine the degree to which they are applicable to the development of high performance models. The software engineering approaches that we believe are appropriate cut across all areas of software development including tools, processes, structure, and standards. These approaches are described below.

- **Establish formal software design procedures**.
  These procedures include formal design documents and periodic design reviews, which facilitate effective collaboration. Software developed using these procedures tends to be well designed, documented, and tested.

- **Modularize, generalize, and standardize code to promote reuse.**
  Good modular design organizes code around the natural data structures, functions, and functional hierarchies implied by a problem domain. Modules are largely self-contained and can often be used without intimate knowledge of their internal detail, and, if well designed, can be useful in multiple contexts. Standardization of module interfaces promotes interoperability. These features promote code reuse that in turn makes efficient use of programmer resources and improves software reliability.

- **Develop layered software**.
  Layering is a way of organizing software on a larger scale than modularization. Lower level software modules are general in function and machine specific in implementation. Higher level modules are specific in function and machine independent in implementation. Layered software promotes an efficient allocation of labor resources; scientists focus on high level applications, while software engineers and computer scientists focus on lower level implementation methods. Code written on top of layered software is more portable, since the machine specifics are localized within a manageable set of software functions.

- **Use software development tools.**
  These include: common code repositories to encourage code sharing; code version control systems that enable collaborative development; issue-tracking tools to collect and preserve information related to codes; automatic configuration tools to facilitate software installation, and automated test suites. Collectively, software development tools increase staff productivity and improve code quality.

- **Object-oriented techniques and frameworks.**
  In object-oriented programming, data, related attributes, and operations for manipulating the data (called methods) are collected into “classes.” A particular instance of a class is called an object. A powerful feature of object-oriented programming is that it hides or “encapsulates” the internal complexity of objects. Frameworks are a promising mechanism to achieve large-scale software integration. Part of the strategy proposed here is to analyze existing frameworks to determine what has and has not worked in related object-oriented software design projects. We will verify these results with high priority pilot projects focused on the needs of the large simulation initiatives underway at NCAR. Based on this analysis, we will decide on the best strategy for the development and use of frameworks.

4.2.2 Model Implementation Issues

THE PLAN: NCAR will ensure that the structure of its large simulation codes is able to efficiently exploit highly parallel, high performance computer systems.

High performance computer systems are changing and becoming more complex, with deeper memory hierarchies and multiple levels of parallelism. Since most of NCAR's applications are severely constrained by computational performance, we must use these systems as effectively as possible. The NCAR simulation environment will adopt a paradigm for software development that simultaneously manages software complexity, scales to large numbers of processors, provides reliable tools, and supports the scientific process (i.e., code is open to peer review, experiments are reproducible, and results are verifiable) while maintaining computational efficiency.

As noted in the 1998 *DoE/NSF Workshop on Advanced Scientific Computing*[10], the challenge is immense:

> *"The general consensus among participants at this workshop was that no simple extension to computing practice as we know it today will carry us to effective use of the next generation of teraflop systems, i.e., systems requiring thousands of processors to achieve sustained multiteraflops capability. Rather, we shall need to solve a wide range of technological problems in qualitatively new ways ... However, the challenges do not all need to be met simultaneously. Judicious selection of key technologies for early development should lead to useable systems that then improve over time in functionality and efficiency. A staged approach, in which we do a good job at each level, is much more likely to be successful than an attempt to solve all problems at once. We must not accept mediocre solutions at any level ..."*

The efforts required to efficiently exploit teraflop systems are more than can be met by any single university department, NCAR division or project, or institution alone. The work must be shared among the larger computational science community, and software development must exploit technologies that promote software reuse to the maximum extent possible. Reuse amortizes development effort across a broader span of effort and resources and contributes to reliability through more thorough review and testing.

4.2.3 Numerical Algorithms and Measures of Suitability

THE PLAN: NCAR will continue to actively define and evaluate measures of suitability for the suite of algorithms required to simulate geophysical systems of interest.

NCAR will continue to conduct algorithmic research and to interact widely in order to stay abreast of relevant algorithmic progress. We will identify gaps that need to be filled and will augment algorithmic efforts at NCAR and/or establish strategic alliances to do so. A major emphasis of this activity will necessarily involve exploration of algorithms that can be implemented efficiently on highly parallel computer systems.

The development and assessment of numerical algorithms for complex simulation systems is a formidable challenge. Ultimately, it requires a subjective judgment of the right balance between solution accuracy and efficiency.

Evaluation of solution accuracy is complicated by its dependence on numerous model and application-specific factors such as model geometry, boundary conditions, the interactions of explicitly resolved motions with parameterized physical processes, and the dominant scales of motion. On the other hand, algorithm efficiency depends on both the characteristics of the particular algorithm and its implementation on a particular machine architecture. Consequently, phenomenological characteristics, machine architecture, and the physical applications of interest all play a role in determining the most appropriate collection of numerical approximations for the realistic simulation of any geophysical system.

There are many examples where the community has identified and evaluated effective numerical approximations for specific scientific problems. For the dynamical cores of global atmospheric models, the shallow water test cases are recognized as fundamental measures of algorithmic suitability. For smaller scale models, a number of idealized flows, such as mountain waves, gravity currents, and warm thermals, have been used as test beds for the evaluation of numerical algorithms. However, our ability to assess quantitatively the suitability of numerical algorithms drops markedly in applications exhibiting additional flow complexity and parameterized physics.

NCAR has traditionally conducted community-based workshops as one means of following emerging research and technology topics, and to engage with current and potential collaborators. We will work to enhance current workshop plans to explore more fully algorithmic progress and development with application to highly parallel computer systems.

4.3 Theme Three: Data Management, Metadata, Post-Processing, and Visualization

Data are the currency of the end-to-end simulation environment. Simulation-based research on terascale computing platforms generates massive data sets with the concomitant difficulties of storage, access, transfer, manipulation, and documentation. The integrity of a research project depends on a data management plan that addresses the passage of data through the end-to-end simulation environment and supports all phases of simulation-based research: hypothesis formulation, boundary and initial condition definition and assembly, model output, post-processing, and visualization. The requirements for each of these phases need to be analyzed and balanced to define the data services that are needed, in addition to meeting simulation and post-processing performance requirements.

Currently at NCAR, data matters are addressed as isolated and disjointed topics; this characteristic is carried forward into projects which themselves are often carried out in isolated and disjointed ways. The problem is further exacerbated at an organizational level because of no coordinated strategy to manage data. The cumulative effects of this situation represent a significant impediment to developing the end-to-end environment in the first place.

However, such an organization-wide infrastructure will ensure long-term commitment by the scientific and technical staff to the concepts to be developed in subsequent sections, and is fundamental to the viability and conduct of the intended services and their supporting software frameworks. As the complexity and resource requirements of very large scientific simulations grow, the resulting datasets will become valuable intellectual resources that may be, in practice, too costly to be reproduced. The datasets themselves must be preserved and openly available, including the tools and methods to access them, and all of these may be of such a nature that they become part of the body of published science

4.3.1 Data Services

THE PLAN: NCAR, working with UOP, will develop a coordinated approach to the management and provision of datasets, and fully implement distributed data transport services.

At the present time, data handling within most simulation projects is a secondary consideration. As data assimilation and data fusion become more routinely used for answering fundamental science questions, internal handling and management of data must become integral parts of the model and code design. Data stewardship, including integrity, validation, and longevity, is an important issue that will be addressed by NCAR and the community.

Another fundamental concept in the end-to-end simulation environment is to make data services uniformly available to all elements of the environment with transparent wide area access to these data services—and thus, provide seamless access to the actual data independent of how and where they are stored. Interpretation of simulation output, particularly when used in conjunction with observational data, often requires reformatting, subsetting, regridding, resampling, and transformation. Tools and strategies must be developed to minimize the overhead associated with these processes and make them transparent to the user. NCAR and UOP data centers may need to provide dedicated processing engines for such purposes. In designing access to data services from applications used in a simulation project, it is necessary to introduce abstractions of the fundamental data structures and define functional interfaces to them—in essence, to develop data services components. This, in turn, will allow appropriate design of data service management within the simulation models themselves and will hide implementation details such as the interface between parallel I/O systems and use of self-describing, platform-independent file formats.

4.3.2 Metadata

THE PLAN: NCAR will collaborate with the community to define and establish metadata conventions and propagate these to the larger standards communities.

Efficient exploitation of massive data sets requires cataloging and documentation through the use of metadata, i.e., data describing the primary data objects themselves. In addition, verifiability of simulation-based research requires systematic collection and maintenance of metadata that document the design and execution of a simulation or collection of simulations.

Locating science information within the massive data archives is currently difficult and requires considerable intimate knowledge of the organization and structure of the archive. To facilitate discovery, metadata must be standardized and organized into databases that support a variety of query types. Different classes of queries require different types of metadata to identify information such as what data are available, the nature of the data, how they were generated, and where they are located.

Current metadata standards used in the community (COARDS, CSM, GDT, SOHO-FITS, CEDAR, etc.) address primarily the description of the contents of individual files. These standards need to be extended to encapsulate information about data collections and their derivation history. For example, environmental simulation systems are often composed as distributed applications; each component can represent a physical subsystem, such as the atmosphere, the ocean, or a level in a grid hierarchy. Each component may be responsible for its own output processing. Metadata must identify the relationships between the components to allow reconstruction of the overall simulation configuration. Similarly, data will often pass through many post-processing steps after the completion of the simulation. Each of these steps needs to be documented in the metadata.

Tools at NCAR to support the tasks of creation of metadata, populating databases with metadata, and querying metadata catalogs are either currently nonexistent or inadequate. Since the actual amount of metadata for different datasets can vary, it will be necessary to allow for different levels of populating metadata. The syntax and semantics of queries on such metadata catalogs must support searches based on science questions in addition to the more conventional content-based queries. It will be necessary to develop a coordinated approach to populating, using, maintaining, and presenting consistent metadata catalogs across the NCAR, UOP, university, and related communities.

4.3.3 Analysis and Visualization Capabilities

THE PLAN: NCAR and its collaborators will develop the means for analysis, visualization, and extraction of information from vast data archives and provide these technologies to the research community

As generators and consumers of the results of terascale computations, the environmental sciences community faces a difficult question: how does one explore and interpret one terabyte, 10 terabytes, or 100+ terabytes of data? To derive any, much less maximum, scientific results from large-scale simulations requires significant and substantially new ways in which data are analyzed and visualized. Analysis of such large data sets poses a significant computational challenge in and of itself. Post-processing and visualization software along with the simulation codes, must be adapted to scalable computing platforms. In addition, many of the low-level operations required by post-processing tools are common to simulation codes, e.g., parallel I/O or stencil operations on fields. Therefore, to the extent possible, component software development should be coordinated between simulation codes and pre- and post-processing codes. Scalability, interoperability, and reliability will thus be enhanced across the project.

The post-processing and visualization requirements in the atmospheric and related sciences community span a range of activities from exploration to explanation. For instance, the dynamics of events may best be explored using volume-rendered visualizations and animation, whereas the dynamics of climate generally requires statistical reduction of the data by definition. Geophysical turbulence requires both types of representations. Exploration with sophisticated visualization is often a first phase, giving insight into the phenomena and leading to the formulation of hypotheses. Complex mathematical manipulation leads to condensation of the data sets to explain the phenomena via simple graphics such as line plots, which prove or disprove the hypotheses. The end-to-end simulation environment is designed to address this diversity and breadth and build upon it. Figure 2 shows a 3-D rendering of clear air turbulence produced in SCD's Visualization Laboratory. Using visualization of simulation data scientists were, for the first time, able to describe and postulate horizontally oriented vortex tubes.

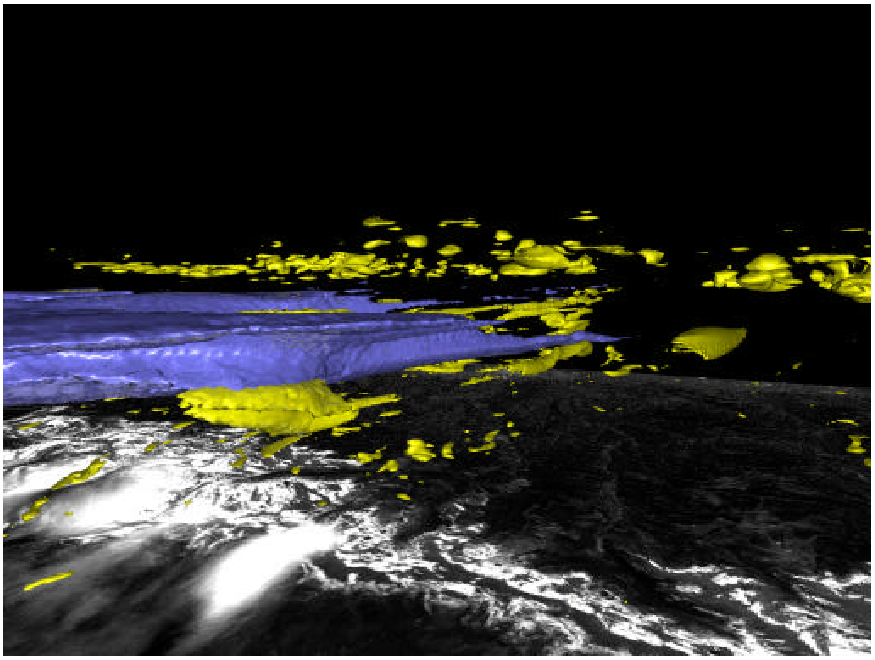

Figure 2: Clear Air Turbulence. Visualization of the jet stream (blue) and fast-moving parcels of air (yellow) over the Colorado Rockies. The view is from the southwest with the Rocky Mountains in white and Boulder, Colorado, in approximately the middle of the image. Courtesy of Don Middleton (SCD) and Terry Clark (MMM).

An accompanying strategy for analysis and visualization tools is their deployment on platforms and operating systems from the laptop through conventional workstations to the highly distributed supercomputer. Assessment of high-end computational capabilities will consider data management, analysis, and visualization portability and performance on a par with the simulation models themselves. Visualization problems such as these are outside the scope of the commercial marketplace, and an extensive research and development program is required to meet future requirements. Visualization systems must be adapted or developed specifically for earth science problem domains. NCAR will complement its simulation capability with specialized, high performance computational resources targeted directly at the distributed visualization and analysis aspect of the research process. NCAR will foster the development of problem-solving environments and frameworks for enabling wide-area, large-scale, collaborative data analysis and visualization for its communities. This will be a highly collaborative endeavor involving other NSF research centers, academia, and DoE efforts such as the Data and Visualization Corridors program.

Terascale simulation data sets will be so large that a significant portion of them may not be able to be moved from their point of computational origin to the desktop. To address the sheer size and inherent complexity in the grand challenge questions, expertise in advanced analysis and visualization techniques and their interrelation are essential. NCAR will develop and deploy new algorithms required to support and extend research efforts in visualization and analysis. For example, new techniques will be required in the form of intelligent agents that are constructed to "mine" large data sets or metadata sets searching for specific features and/or events. These agents would adapt their region of inspection according to what they are taught, and then visualize or represent their findings.

4.4 Theme Four: Project Organization and Management

THE PLAN: NCAR will organize the composition and management of its large modeling efforts to maximize the effectiveness of multidisciplinary teams.

For the past two decades, major models at NCAR have been developed primarily by scientists and programmers in the science divisions, while the scientific computing division has primarily provided computing hardware, services, and consulting advice. The complexity of both the simulation models and the computer systems they run on continue to increase. In addition, community models need to be better integrated into broader national initiatives. Together, these three realities require new organizational paradigms, which must transcend disciplinary boundaries and organizational structure. Thus software development activities require a new blend of skills, training, approaches, and energies. For projects of this kind, we face new challenges in finding approaches that promote and reward broad participation, manage diverse resources efficiently, and maintain an uncompromising focus on science objectives and excellence.

4.4.1 Team Collaboration in Model Development

THE PLAN: NCAR will ensure that the organizational structure of its major modeling projects integrates the expertise of both physical and computer scientists.

A successful team effort in developing major modeling systems requires institutional recognition of and commitment to the projects and their team members. In forming project teams, NCAR will identify the roles that are needed to achieve the project goals, roles that should be addressed within NCAR, and roles that are more appropriately filled by establishing collaborations and partnerships. Teams must be assembled and integrated as early as is feasible so that there is an ongoing relationship among the team members; each member must be committed and accountable to the project, and must share in the rewards for the project's success. In the context of this strategic plan, this means that teams of physical scientists, computer scientists, computational scientists, and software engineers need to be integrated from the beginning of model development projects, and be dedicated to achieving both the scientific and computational research objectives of the project.

Bringing team members from diverse disciplines to participate in all phases of the project entails an educational process in which specialists in each scientific area gain appreciation for the perspectives of collaborators in other areas; the resulting design and implementation strategy benefits from this cross fertilization. In such a mode, a critical path for the project can be established, agreed to early on by all parties, and adjusted as needed, always within the context of the project's research goals. From a personnel perspective, it is important for team members to have well-defined responsibilities and accountability, and to ensure that professional development opportunities exist equally for all team members.

4.4.2 Project Management and Resources

THE PLAN: Management of major software projects will accommodate the broad constituency of the team effort, facilitated by project managers who have authority commensurate with their responsibility.

Creative management practices and constructs are required to direct major software projects involving diverse teams of specialists spanning NCAR divisions, UOP programs, universities, and national laboratories and agencies. To be successful, project management should include mechanisms for project oversight by those responsible for providing resources to the project, for advisory guidance from the broader community, and for direction and coordination of the project activities. Accommodating these functions requires formal management structure that is supportive of and responsive to the science while promoting requisite cohesion, rewards for participants, collegiality, and accountability. At each level, appropriate and proper authority must be vested in the managers of these software projects if they are to be able to accomplish institutional goals and meet their responsibilities as managers.

THE PLAN: NCAR and UCAR will develop a coordinated strategy for developing simulation models that allows the institution to be agile in allocating resources for the development and evolution of all aspects of the end-to-end simulation environment.

Institutional resource management for large simulation activities currently is constrained by the sources of funding. The acquisition and deployment of fiscal resources needed to support these projects present formidable challenges for managers and scientists alike. While PI-like funded activities and large projects complement each other, accountability to the mission-oriented funding agencies often limits participation by NCAR in large collaborative community-oriented projects. Consequently, an adequate level of base support for major modeling projects should be available to ensure the continuity of these high-priority efforts. As in promoting any activity successfully, NCAR and UCAR's commitment to obtaining, allocating, and supporting simulation projects and goals in an ongoing way must be visible and genuine.

4.4.3 Community Support

THE PLAN: NCAR will provide adequate resources to support the necessary model documentation, distribution, and training required for effective community use of major modeling systems.

The science groups responsible for development of major NCAR models historically have provided support for these community models. Although the commitment to share these modeling resources with the community has always been strong, resources to support these activities have often been inadequate and only obtained in an *ad hoc* manner. These community-wide models have been described as facilities in the UCAR system, and will increasingly be regarded as such. Accordingly, they require the appropriate staff, funding, and organizational commitment to ensure that the community receives the support it needs.

4.5 Theme Five: Computing Profession at NCAR

4.5.1 Role of Computer Science Research at NCAR

THE PLAN: NCAR will embrace computer science expertise by increasing its emphasis on computer science related research and actively integrating this research into simulation projects.

An increase in and reorganization of computer and computational science efforts within NCAR will be needed in order to fully integrate computer specialists into large modeling efforts and to realize the end-to-end simulation environment envisioned here. Recognizing the reality of resource limitations, NCAR cannot, nor should it, simply broaden its mission to encompass the breadth of computer and computational science research; rather, we must augment our computing related activities in strategic areas to foster the simulation environment needed by our scientific and technical constituency. We need also to strengthen ties to the larger computer science community, forming partnerships to jointly pursue major computer science oriented projects of mutual interest. Areas of priority should include research and development in software frameworks, numerical algorithms and libraries, model building blocks, rapid prototyping, database management, metadata conventions, and post-processing analysis and visualization tools. NCAR should be recognized as a place where computing professionals can establish and maintain rewarding careers.

4.5.2 Role of Software Engineering at NCAR

THE PLAN: NCAR will provide opportunities for software engineers to explore and employ state-of-the-art techniques and methodologies in challenging simulation projects.

Software engineers play a critical role in the design, implementation, and testing of software they employ, especially for large or complex projects. They must employ state-of-the-art methodologies and take maximum advantage of the best available techniques. Thus, it is important for individuals filling these roles to participate actively in their peer community and have access to continuing education. Attention must be given to all aspects of software engineering, providing suitable tools, time, and resources for documentation, design reviews, and the development of new techniques.

It is desirable to preserve the role of senior software engineers for advanced level design, technical project management, and programming, with less emphasis, where possible, on personnel management. UCAR should sponsor forums for the exchange of ideas, practices, and common problems and their solutions among software engineers from across the entire organization. Software engineers should be assigned projects that challenge their capabilities and provide opportunity for professional growth.

4.5.3 Integrated Teams for Simulation Projects

THE PLAN: NCAR will promote a broader use of joint appointments, particularly between the computing and science divisions, to strengthen the recognition of commitments to major team development efforts.

To achieve success in recognizing and rewarding computer science contributions within the organization, NCAR must create an environment that both stimulates cutting-edge research in computer and computational science and integrates this research with large simulation projects. For example, one tool for encouraging and recognizing the institution's commitment to multidivisional, multidisciplinary projects is wider use of joint appointments. This technique has been successfully exploited among the Mesoscale and Microscale Meteorology and Atmospheric Technology Divisions and the Research Applications Program, and elsewhere in several instances. While not in and of themselves a solution that ensures the success of team efforts, joint appointments provide visibility and recognition of long-term commitments for interdivisional interactions. Joint appointments also aid in maintaining accountability to projects through the joint responsibility of managers in different organizational units for performance and project reviews. A supporting environment for interdivisional research will, over time, allow expertise to migrate where it is most needed.

4.5.4 A Refocused View of Computing Professionals

THE PLAN: NCAR, recognizing the increased emphasis on computational research, will increase the use of the scientist category for computing professionals.

We must ensure that UCAR's personnel policies provide the appropriate recognition for contributions of computing professionals. Success in achieving the vision in this plan will require hiring additional computational and computer scientists to establish a credible research emphasis. To facilitate these efforts, NCAR will foster the creation of a critical mass of computational professionals in the scientist appointment ladders and will promote appropriate research collaboration with the broader research community. Appointment as a scientist at NCAR carries with it a priority and responsibility to conduct research that leads to peer-reviewed journal publications. The important role of computer specialists in multidisciplinary teams and computational research warrants investment in seminars, training, attendance at national and international meetings, and encouragement to develop collaborations with peers within NCAR and other organizations, including short- and long-term visiting appointments. In support of this recommendation, trends and practices at other government labs such as the Advanced Computing Laboratory at LANL, the Math and Computer Sciences division at Argonne National Laboratory (ANL) and the High Performance Computing Division at National Energy Research Scientific Computing Center (NERSC) should be examined along with private industry and university models.

THE PLAN: NCAR will re-examine the position descriptions in technical job categories to ensure that they are consistent with the roles needed in these interdisciplinary team projects.

Finally, with increasing emphasis on integrating computational and computer scientists into multidisciplinary projects, position descriptions for these job categories may need to to be revised to reflect their expanding roles within the organization. This re-examination will help affirm the value of modern computational professionals to the organization, and strengthen the perception that NCAR is an attractive place for them to make professional careers as research peers. Furthermore, we must ensure that the UCAR compensation structure is conducive to recruiting and retaining senior staff and, at the same time, is equitable to all multidisciplinary team job categories.

4.6 Theme Six: Collaborating in a Distributed Environment

To succeed in providing the capabilities outlined above, UCAR faces a number of challenges. At present, much of the effort and resources that support work in a distributed environment are in individual research projects and applied in an *ad hoc* manner to address specific needs. Significant additional resources and new roles for participants are required to establish effective collaboration tools that can be readily used throughout the community. Methods of support must have configurable levels of capability on the desktop (e.g., sound, white board, visualization, remote desktop access) that are tailored to the needs of a broad diverse community of scientists and educators.

4.6.1 Distributed Access to Facilities and Resources

THE PLAN: NCAR will adopt and build upon the "collaboratory" concept: supporting scientific collaboration; exchanging information; sharing and debugging code; planning experiments; finding, sharing and analyzing data; and providing consulting and training, all within a distributed environment.

In the first half of NCAR's 40 years, accessing NCAR resources, expertise, and services required that members of the computing community travel to Boulder to avail themselves of NCAR facilities. Beginning in the early 1980s, users could remotely log in to an NCAR computer, but processing was slow, awkward, and expensive, and the amount of data that could be transferred back and forth was extremely small. New and emerging information technologies now provide easier remote access to NCAR's facilities and resources, both computational and personnel. Providing access via Web servers is just the beginning and ultimately may not be sufficient to achieve the levels of functionality required. Significant opportunities exist for NCAR to learn more: for example, digital libraries are presenting new paradigms for how people find and exchange information. The Upper Atmosphere Research Collaboratory and its successor, the Space Physics and Aeronomy Research Collaboratory, have broken new ground in bringing these concepts to a demanding research community to support real-time instrument control, planning of observing programs, data presentation, and investigator interaction. NCAR's challenge is to build upon instances such as these to provide an effective collaboratory environment for its broad multidisciplinary constituency.

THE PLAN: NCAR will adopt the concept of location independence for its facilities and resources.

The current nature of facilities and resources is usually tied to the specific location in which they reside. In a distributed environment, users become overwhelmed by having to learn unique aspects of particular environments. Thus, in first adopting and then extending the concept of location independence, it will be necessary to expand beyond the present views of facilities and resources and aim for location transparency, i.e., a common method for access and use. This concept is familiar to many people in the form of the Uniform Resource Locator (URL) for accessing Web pages and resources, and its lesser-known generalization the Uniform Resource Name (URN). A URL can hide details of location and even organization of information; the concept applies equally to high performance computing resources and services. NCAR must fully embrace these architectural considerations in as many of its software systems as resources allow, and must use them to coordinate its diverse computational facilities, services, and resources in support of a distributed environment.

4.6.2 Tools for Collaboration

THE PLAN: NCAR will research, develop, and implement the infrastructure to facilitate working in a geographically distributed environment.

Collaborators in our distributed environment will need to pursue and invent new ways to share data, models, and ideas across the country and around the globe as readily as they do with their colleagues across the hall. In recent years, the Web browser on a desktop system has become a common window to a variety of information sources, but as such is a very limited collaboration tool. Applications that run on the desktop often require substantial system resources and are complex to learn and remember are at the other end of the spectrum. Advances in the understanding of how to use multitiered architectures to bridge the gap between ease-of-use and adequate functionality are just beginning to appear. A basic need is that tools must have common operational principles, like a browser, and not require the user to relearn these principles each time. Many lessons can be learned from the application of techniques developed for distance learning and those being explored in digital library projects.

If suitably specified, novel collaboration tools will be the source of important advances in interdisciplinary science in the near future. Therefore, every effort must be made to ensure that resources are allocated for migrating tools to different environments and maintaining them; that the tools are chosen with end use as the primary factor, rather than being technology driven; and that tools can scale from developer to end user and from laptop to supercomputer. Additionally, tools for the distributed environment must be integrated with model development, application tools (commercial and otherwise), analysis packages, and data management. Consequently they should contribute to the design of, and leverage, the component frameworks discussed earlier.

The challenges in this area are not unique to NCAR and we will, wherever possible, use off-the-shelf technology for these tools.

4.6.3 Ongoing Commitment

THE PLAN: NCAR will continue to participate in leading-edge research projects with strong community participation to effectively build upon the current generation of collaborative techniques and tools.

Investments in tools for collaboration and integrating geographically distributed project teams are only the first part of a long term NCAR commitment. In essence, what is needed is a long-term commitment to accommodate multiple generations of project, technology, and personnel changes. To optimize these investments, it will be necessary to develop metrics for evaluating how well strategies and their implementations are succeeding and where they need revision. Many institutions are addressing these issues, specifically in the context of the Web, where reports showing "hits" of Web pages, active and multi-user sessions, session duration, and number of bytes transferred only scratch the surface in assessing the effectiveness a particular site.

As is evident from many successful community projects, NCAR must establish and maintain solid multimedia documentation from a common source to facilitate community interaction, provide education and training materials, and guide future development. NCAR will continue to assess balances among the many types of interactions required to support different participants in simulation projects, as well as to define the level of support that can be made available to them. Finally, since many of the collaboration requirements expressed herein are not unique to simulation projects, appropriate attention must be given to broader efforts within NCAR and UCAR and the community.

5 Concluding Remarks

Achievement of the vision presented herein will require that substantial changes to the methodologies of large-scale simulations at NCAR and in the university community be adopted. Creation of an end-to-end simulation environment will involve new hardware, new software development approaches using interdisciplinary teams, and adequately scaled data storage along with rapid and efficient access, analysis software, and innovative visualization techniques.

The problems faced are scientifically and technically demanding, and new development and research paradigms will alter the social fabric of the people in our scientific community. The implementation of the necessary changes will present significant challenges to management.

While this document has been written as a strategic plan for NCAR, it is also intended to reflect the goals for the broader research community. This whole community must be substantively engaged in order for the plan to be successful. The six strategic themes represent an implementation framework. The critical next steps will be to begin this implementation in close collaboration with NSF, the universities, and other national laboratories. Indeed, progress in the atmospheric and related sciences that depend on simulation will be determined, to a significant degree, by how well we are able to achieve the goals presented in this document. We look forward to this challenge and are confident that our strategic plan for large-scale simulation will pave the way for an exciting and rewarding future for our community and that attendant important benefits to society will occur as a result.

References

[1] *UCAR 2001, A Strategic Outlook for the University Corporation for Atmospheric Research*, UCAR, May 1992

[2] *Grand Challenges: High Performance Computing and Communications.* A Report by the Committee on Physical, Mathematical, and Engineering Sciences. Federal Coordination Council for Science, Engineering, and Technology. Office of Science and Technology Policy, the FY 1992 U.S. Research and Development Program, 1992

[3] *From Desktop to Teraflop: Exploiting the U.S. Lead in High Performance Computing*, NSF Blue Ribbon Panel on High Performance Computing, National Science Foundation 1993.

[4] *Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure*, Committee to Study High Performance Computing and Communications: Status of a Major Initiative, National Research Council, 1995

[5] *Computing, Information, and Communications: Technologies for the 21st Century*. National Science and Technology Council. Committee on Computing, Information, and Communications. Supplement to the President's FY 1998 Budget, 1998

[6] *Information Technology Research: Investing in Our Future*. President's Information Technology Advisory Committee Report to the President, 1999

[7] *Geosciences Beyond 2000: Understanding and Predicting Earth's Environment and Habitability*, NSF Geosciences Strategic Plan, NSF, 2000

[8] *Environmental Science and Engineering for the 21st Century*, NSB Interim Report. NSF, 2000

[9] *UCAR Information Technology Council Strategic Plan*, University Corporation for Atmospheric Research, *(http://www.fin.ucar.edu/itc/ITCStrategicPlan.html)*, 1998

[10] *National Workshop On Advanced Scientific Computation*, Department of Energy and National Science Foundation, National Academy of Sciences, Washington, D.C., July 30-31, 1998

Glossary

ANL—Argonne National Laboratory (http://www.anl.gov/)

Class—Defines an object in computer code

Client/server architecture—Specific case of a tiered architecture

COARDS—Cooperative Ocean-Atmosphere Research Data Service (http://www.cdc.noaa.gov/~jac/coop/)

COMET— Cooperative Program for Operational Meteorology, Education and Training

Components—Specific instances of classes that perform a specific function and are designed in a way to easily operate with other components and applications (in the Web context they are referred to as applets). They enable a modular implementation of a specific algorithm intended for reuse.

Computational science—The specific application of computer, mathematical, and discipline science to solve or study problems using computational means

Computer science—The study of computers and their operation, including both hardware and software design, and many subdisciplines

CCM—Community Climate Model (http://www.cgd.ucar.edu/cms/ccm3/)

CSEM—Comprehensive Space Environment Model

CSM—Climate System Model (http://www.cgd.ucar.edu/csm/)

DLESE— Digital Library for Earth System Science Education

DOD—Department of Defense

DODS—Distributed Oceanographic Data System (http://www.unidata.ucar.edu/packages/dods/)

DoE—Department of Energy (http://www.doe.gov/)

Domain analysis—The process of examining a specific problem domain, modeling it to define a project

Encapsulate—To hide information behind an interface, within an object or class

Framework—A family of classes that can be used to compose new programs in an application domain (see also Appendix B)

GDT—Standard defining a set of conventions to promote the interchange and sharing of files created with netCDF, particularly for climate and related data (http://www.pcmdi.llnl.gov/drach/GDT_convention.html)

Gigabyte —10^9 bytes

Gigaflops—10^9 floating point operations per second for a computational process on a computer

GLOBUS—Project to develop basic software infrastructure for computations that integrate geographically distributed computational and information resources (http://www.globus.org/)

Grid—A heterogeneous, geographically distributed complex of advanced computers, networks, storage and display devices, and instruments.

Hierarchies—Nested grouping of object classes with greatest commonality highest in the hierarchy

Interfaces—Definition of what information and protocols may be exchanged between tiers

IT—Information Technology

ITC—Information Technology Council

Java—A platform-independent, object-oriented language developed by Sun Microsystems (http://java.sun.com/)

JOSS—Joint Office for Science Support

LANL—Los Alamos National Laboratory (http://www.lanl.gov/)

Layering—A way of organizing software on a scale larger than modularization and encapsulation, promotes efficient allocation of labor resources in developing new codes and maintaining existing ones

LDHwg—The Large Data Handling working group

MMM— Mesoscale and Microscale Meteorology Division

Metadata—Information that accompanies data to provide guides to structure, content, and context

Middle tier—Also known as service layer, component layer, middleware

Modeling—Use of object-oriented techniques to understand project/model resources, processes and organization

MSS—Mass storage system

Multimedia—Use of computers to present text, graphics, video, animation, and sound in an integrated way

NASA—National Aeronautical and Space Administration (http://www.nasa.gov/)

NASA/DAO—National Aeronautical and Space Administration/Data Assimilation Office

NCAR—National Center for Atmospheric Research

NCSA—National Computational Science Alliance at the University of Illinois, Urbana-Champaign (http://www.ncsa.uiuc.edu/Indices/InsideNCSA/NCSAbrief.html)

NERSC—National Energy Research Scientific Computing Center (http://www.nersc.gov/)

netCDF—The network Common Data Format (http://www.unidata.ucar.edu/packages/netcdf/)

NOAA—National Oceananic and Atmospheric Administration (http://www.noaa.gov/)

NPACI—National Partnership for Advanced Computational Infrastructure at the University of California in San Diego (http://www.sdsc.edu/)

NSF—National Science Foundation

Object—Represents a real or abstract thing with a name

Object-oriented—Develops a definition of some process or entity considering all attributes

Open standards—Written guidelines for the development of computer systems that are not owned, licensed or restricted to one organization or entity. Systems developed using open standards are able to interoperate with other systems that conform to the standards

ORNL—Oak Ridge National Laboratory

PACI—Partnership for Advanced Computational Infrastructure, (http://www.interact.nsf.gov/cise/html.nsf/html/paci?OpenDocument)

PCM—Parallel Climate Model

Petabyte—10^{15} bytes

Petaflops—10^{15} floating point operations per second for a computational process on a computer

SCD—Scientific Computing Division, within NCAR

Software engineer—An engineer or programmer who designs, develops, and manages computer codes

Software engineering—The computer science discipline concerned with the field of code development.

Terabyte—10^{12} bytes

Teraflops—10^{12} floating point operations per second for a computational process on a computer

Terascale—A collection of computing resources (processor, storage, I/O bandwidth, and networking) which are about three orders of magnitude greater than that found on conventional departmental servers

Three-tier architecture—Definition of a separation between data (lower tier), common services (middle tier), and the user (top tier)

UARC—Upper Atmosphere Research Corporation

UCAR—University Corporation for Atmospheric Research

UNAVCO—University NAVSTAR Consortium

UNIDATA—University Data Interactive Computing and Communications Systems

UOP—UCAR Office of Programs

URL—Uniform Resource Locator

URN—Uniform Resource Name

Web—World Wide Web

WRF—Weather Research Forecast Model (http://www.mmm.ucar.edu/wrf/)

Appendix A: UCAR Information Technology Strategic Plan Summary

UCAR Information Technology Council
Boulder Colorado
September 1997–September 1998

UCAR's vision for its information management activity is to develop a corporate-wide information strategy that reflects UCAR's six goals, fosters information sharing effectively and economically across UCAR and the communities it serves, and employs technologies that improve the corporation's efficiency.

The strategic direction recommended for UCAR is to create an information environment that ultimately can support all-electronic operations, effective communications (corporate-wide and external), departmental specialization, and archiving, all with appropriate levels of security. Intellectual artifacts (e.g. papers, software, equipment, ideas, etc.) are among the most important resources and outcomes associated with UCAR's mission and efforts, and this architecture recognizes that such artifacts, along with financial transactions and business documents, are becoming entirely digital at a strikingly rapid rate.

UCAR's strategy for achieving this vision is embodied in eight specific recommendations:

1. UCAR IT should move toward distributed, object-oriented methodologies

   These methods should be accompanied by object modeling on an enterprise-wide scale. Anticipated benefits include: more rapid and robust software development; software systems which better match UCAR's needs and which are easily adapted as those needs change; greater reuse of existing software components; and better utilization of the distributed computing capabilities provided by our networks.

2. Software systems in UCAR should be designed around a three-tier architecture, with emphasis on standardized interfaces in the middle (service) tier

   This architecture allows the separation of both user-interface and data-storage components from the service-provision layer. Hence legacy databases can be encapsulated and made available through the net, and users can more easily perform a wide variety of daily tasks with their favorite applications. A service tier also will facilitate implementing UCAR's business rules and common services (policies, overhead rates, work flows, authentication, etc.) in an appropriately centralized fashion.

3. UCAR should continue to operate and support a leading-edge network and associated services

   A top-quality network is required infrastructure for IT advances and is increasingly essential for UCAR to fully utilize its intellectual resources. This network must support multi-media applications, such as video conferencing, as well as more traditional data flows. Furthermore, the net is becoming UCAR's primary interface to the external world, ranging from basic Web pages and data access, to complex scientific collaborations and distributed computations.

4. UCAR should employ open standards whenever feasible

   Open standards provide numerous advantages over proprietary (vendor-specific) standards: implementation choices are increased, multi-vendor hardware and software interoperability is enhanced, and the risk of dependency on a single vendor is reduced. Open standards (including TCP/IP, the World Wide Web, etc.) historically have thrived much longer than proprietary ones.

5. Continued innovation at the divisional/program level should be encouraged and supported

   Technological progress at UCAR often has been driven by individual programs striving to fulfill their missions. While common standards must be offered and their use encouraged, individual groups will remain free to solve problems in ways they judge to be best. A suitable implementation framework must be put in place to enable groups to take advantage of institution-wide advances as well as to leverage efforts at the division and program level, where warranted.

6. Continued emphasis must be placed on technical training for staff

   UCAR can exert leadership in the IT arena only if staff are conversant with current technologies and have appropriate tools. The rate of change in these areas requires continual learning. Without adequate and ongoing training, adoption of new tools and techniques is unlikely to be successful.

7. A standing committee should be formed to coordinate IT efforts in UCAR

   This group should be based on the existing IT Council and include liaisons with the other IT-related committees (DSAC, CSAC, NCAB, etc.), with careful selection from UCAR's research, technical, and managerial staff. This group would: further refine the IT strategy; oversee the development of service layer (and other) standards; and coordinate UCAR's various IT efforts. The committee should report directly to the UCAR President.

8. UCAR should consider creating new funded engineering groups to design and implement essential service-layer functions

   Key services—such as authentication—which must be developed and administered centrally, would be handled by such groups. Additionally, responsibility must be assumed for: refining the enterprise-wide object model; developing, disseminating, and maintaining a library of reusable classes and other software to support the model; and addressing other IT-related issues that concern the entire institution, such as e-mail with attachments or tools for network conferencing and collaboration.

   UCAR also should consider appropriate administrative structures for the new engineering groups. While a single administrative entity may eventually be desirable for these groups, the present recommendation is to allow a distributed management structure as indicated in a later section. This compromise represents a practical approach to allow existing engineering groups to continue operating without any significant disruption and at the same time, to develop new groups supported by divisions or programs where needs or opportunities arise.

The vision and strategy were developed through UCAR-wide workshops, and by modeling groups who developed case studies of UCAR's varied information dissemination, communication and manipulation needs by functional area (research, education, management etc). These functional sets were derived from preparing a model of how UCAR shares and uses information - research, technical and business - both internally and externally. All of these discussions led, in addition to the specific recommendations, to the very important general conclusion that the handling of these multiple 'kinds' of information has significant commonality - this is not to say the information is the same, but rather that the paths of use and transformation in fact are. The importance of this general conclusion is that it provides the rationale for recommending object-oriented methodologies and the three-tiered architecture.

With respect to implementation, UCAR's strategy will be to build incrementally on many efforts across the institution that are consonant with the IT recommendations. This fashion of implementation leverages resources throughout the institution and is likely to be the lowest cost approach. For example, networking as managed in NCAR's Scientific Computing Division, and the plans in UCAR's Finance and Administration group, to examine and meet business needs systematically and in a standard way will be encouraged, expanded where possible, and promoted as examples of better ways to do business.

In times of rapid IT advances, an additional important aspect of UCAR's implementation effort will be to identify and resolve the most pressing 'high return' issues first. For example, there is widespread interest, expressed in the workshops, modeling sessions, and e-mail communications from staff, for solving some problems quickly; dominant issues seem to be: transfer of attachments via e-mail, 'smart' forms creation, authentication systems, and embedding of business rules within computerized procedures. There is also widespread agreement that a robust, ongoing training program for staff is essential.

The first half of the UCAR Information Technology strategic plan describes the rationale for its development, UCAR's Information Technology vision and its attributes, and the ITC's eight specific recommendations. For those readers interested in the underlying technologies, the second half of the plan elaborates upon those strategies; describes the modeling activities, the system architecture and implementation approach, and the strategy for immediately advancing IT at UCAR, specific action items and goals for the next two years, and some conclusions.

Appendix B: Object-Oriented Code and Frameworks

Object-Oriented and Object-Based Code

In object-oriented programming, data, related attributes, and operations for manipulating the data (called methods) are collected into "classes." A particular instance of a class is called an object. For example, a vector class might include a class attribute of length, and a class method called get_length. An object of that class might be a vector with a length of 5. To access the length of the vector, the vector method get_length would need to be called. A powerful feature of object-oriented programming is that it hides or "encapsulates" the internal complexity of objects. The programmer using a vector object would not need to know how its data were laid out in memory, or how its length was obtained. Thus, object-oriented software design organizes data and methods into functional components to make the process of developing and maintaining large, complex software systems more manageable.

Object-based programming is a looser interpretation of object-oriented programming and may include some procedural elements for high performance or convenience. An object-based implementation of the vector class described above might access the vector length by de-referencing the vector class structure directly, without calling a class method.

Object-based and object-oriented design methodologies have been used extensively over the past decade in fluid dynamics on parallel platforms. Studies show that the object-based approach can separate the numerical scheme from parallelization procedures, which helps to increase portability and readability. This can be done while still attaining good performance. Object-based code in fluid dynamics applications can improve reliability, encapsulate optimizations, and facilitate load balancing in adaptive grid computations and mesoscale modeling. Object-based software has also been used successfully in other computational areas pertinent to earth science models, such as sparse matrix computations.

Frameworks

An object-oriented framework is a family of classes that can be used to rapidly compose new programs in a particular application domain. Like object-oriented code, frameworks help to organize complex codes, though at a larger scale. They provide a way of sharing not just a single class, but a collection of integrated classes. Frameworks provide standard interfaces that enable existing components to be combined and reused; they provide a standard means for components to deal with functions such as error handling, exchanging data, and invoking operations on each other.

Frameworks are often structured as layers of classes. Basic classes for scientific computing typically include a set of simple data structures, communication routines, math utilities, and I/O utilities. These are built as generally as possible without compromising performance (i.e., select critical sections may be written procedurally). A collection of these classes may be modeled as a bottom layer, which can be used to build another layer of more complicated classes. Multiple layers may be supported. Since vendor- or architecture-specific code can usually be restricted to the lowest levels, layering helps to isolate these details from the majority of model developers and users. Another benefit of frameworks is that they foster a uniformity of design that can reduce development and maintenance costs since programmers can more easily adapt their applications to a particular machine architecture.

Frameworks have been classified into “called” and “calling” categories. “Calling” frameworks dictate the flow of control in an application, controlling and invoking a sequence of methods. “Called” frameworks provide interrelated methods or libraries of methods that can be invoked by an application. One can think of a “called” framework as being composed of multiple interoperable libraries.

Some popular examples of layered frameworks are POOMA for general scientific computation and Overture for solving partial differential equations, both of which are designed for high performance on parallel platforms.

Investigation of layered frameworks is essential since layering can be employed to simplify code development and to help insulate the scientific efforts from the volatility in the high performance computing market.

Appendix C: Community Models and Associated Datasets

NCAR has a number of research models that are shared with scientific colleagues in the broader community. Two of these models, the Community Climate System Model (CCSM) and the NCAR/Pennsylvania State University (PSU) Mesoscale Model (MM5), are used widely and have established formal mechanisms for model distribution and support. A brief summary of the community involvement with Community Climate Model (the atmospheric component of the CCSM) and MM5 is provided here.

NCAR/PSU Mesoscale Model: Community Use and Participation

- 600 estimated active users
- 220 user institutions
  - 58 U.S. universities
  - 37 federal and state agencies
  - 91 non-U.S. institutions (30 countries)
  - 34 private companies
- Over 500 refereed publications utilizing MM4 and MM5 during the past decade
- 24 institutions running MM5 in real time (18 in US, 6 non U.S.)
- Annual MM5 users' workshop (60–100 participants per year)
- Biannual MM5 tutorial classes (30–50 per class)

NCAR Community Climate Model: Community Use and Participation

- 300 estimated active users
- 130 user institutions
  - 47 U.S. universities
  - 8 federal and state agencies
  - 51 non-U.S. institutions (27 countries)
  - 25 private companies
- Over 300 refereed publications utilizing the CCM during the past decade
- Annual CCSM workshop (~50 active participants in the Atmospheric Model Working Group)

- Periodic tutorial activities
    - CCM Tutorial Workshops
    - NATO Advanced Study Institute (see *Numerical Modeling of the Global Atmosphere*, edited by Prof. Alan O'Neill, Director Castelvecchio Pascoli, Tuscany, Italy, Kluwer Academic Publishers, May 25 - June 5, 1998)

CCSM and MM5 Datasets:

In addition to large simulation tools, NCAR research efforts make simulation datasets available to the broader user community. Two examples of these datasets, and statistics on their external use, are also described.

Condensed datasets from CCSM control simulations are routinely made available to the community via the Web. Statistics for the first four months of calendar year 2000 show that there have been 2492 downloads of these data files, averaging over 20 CSM data files downloaded per day. The average data volume is approximately 1.7 Gbyte per day, where the data has been downloaded to 57 different domains, which can be broken down by the following domain types:

24% educational
24% commercial
14% non-U.S.
5% other government labs
33% unidentified

The MM5 Modeling group at NCAR runs real-time MM5 weather forecasts and continuously updates the model forecast products on the Web (http://rain.mmm.ucar.edu/mm5/). A global version of the model produces one ten-day forecast daily, a version running over the continental U.S. produces a 48-hour forecast every 12 hours, and a high-resolution nested grid over a portion of the western U.S. produces a 36-hour forecast once a day. For the first four months of calendar year 2000, the Web site for MM5 real-time forecast products has had 20,758 unique visits from 4800 unique hosts. These visits are distributed among the following types of domains:

27% educational
23% commercial
24% non-U.S.
5% other government labs
21% unidentified