Lawrence Livermore National Laboratory – University of California, San Diego
Scientific Data Management

January 14, 2005

Reagan W. Moore (San Diego Supercomputer Center)
Tim Barnett (Scripps Institution of Oceanography)
Michael Norman (University of California, San Diego)

1. Proposal Summary: LUSciD – LLNL UCSD Scientific Data

The University of California, San Diego, proposes a joint collaboration with the Lawrence Livermore National Laboratory (LLNL) on the application of advanced scientific data management technologies. The goal of the project is to improve the conduct of science through the provision of scientific data management technology that enables the organization, manipulation, and analysis of observational and simulation data.

The project is driven by two exemplary scientific applications:

1) Global climate modeling to determine the impact of climate changes on water supply, led by Dr. Tim Barnett. A Detection and Attribution analysis will be used to answer the question: Can we detect a global warming signal in main hydrological features of the western United States? This will involve making runs of global climate and downscaling models that will be unprecedented in scope.

2) Cosmology simulations of the early universe and the generation of simulated observations for the Large aperture Synoptic Survey Telescope, led by Dr. Michael Norman.

UCSD will provide the driving applications for climate modeling and cosmology. The simulation codes will be based on current state-of-the-art algorithms (an experimental version of the CCSM3 climate model and ENZO cosmology code). Regional models will be forced by the output of the global model simulations to improve resolution for assessing impact on water supplies. Adaptive mesh refinement extensions will be used to model galaxies. Both applications will be capable of generating hundreds of terabytes to petabytes of simulation output.

The simulation codes will be executed on compute resources at LLNL, with the output cached on storage systems at LLNL and stored on archives at SDSC. Data grid technology will be deployed between LLNL and SDSC to facilitate the movement of data to UCSD for analysis of by researchers. Existing data collections at UCSD will be federated with the simulation output to enable comparisons with observation data, and support collaborations with researchers within the earth science and astrophysics disciplines. SDSC will research, implement, and provide support for data grid technology that improves the ability to organize, share, publish, visualize, and analyze both simulation output and observational data.

Data management research activities will be selected that improve the ability to conduct science research. Example research areas include:

- integration of the SRB data grid with data repositories at LLNL
- possible federation with the Earth Science Grid for access to the PCMDI collection
- development of data movement capabilities for terascale applications
- integration with disk caches for enhanced analysis of simulation output
- integration with visualization systems for display of hundred-terabyte 4D data sets
- integration with feature detection systems for automating analysis of simulations
- integration with dataflow management system for managing interactions with data
- integration with digital libraries for managing descriptions of features in the data
- demonstration of prototype LSST data management pipeline
- development of terascale database applications

Collaborations will be sought that facilitate interactions with other researchers within the scientific disciplines, broaden the scope to national-scale projects, and promote on-

going research past the initial three-year period. Opportunities to support larger-scale projects such as the Large aperture Synoptic Survey Telescope will be pursued.

2. Project Description

Current simulation codes have the capability of generating hundreds of terabytes of simulation data, comprising millions of files. The ability to organize, manage, and analyze these massive data sets requires the use of advanced data management systems. A major research activity at the San Diego Supercomputer Center has been the development of data grid technology for the automation of all aspects of data management. The technology development has been driven by the data management requirements of application codes in Astrophysics and Earth System Science. We propose a collaboration between LLNL and UCSD on the development of scientific data management systems that can organize, manage, and analyze these massive data collections.

The collaboration will be based on the data management requirements for the ENZO cosmology code and the high-resolution version of the experimental CCSM3 with the finite-volume atmospheric code. These simulation codes will be executed on LLNL compute platforms, the output will be stored on SDSC storage systems, and the data will be analyzed at UCSD through data grid access mechanisms.

2.1 Global Climate Change:

Substantial changes have been observed in the snow pack over the western US in the last 40 years (Hamlet, et al, 2002; Mote, et al, 2004). These changes have led to observed changes in the hydrographs of virtually all western rivers and streams, i.e. the maximum spring peak flow is coming between 2-3 weeks earlier in the annual cycle over this period (Stewart, et al, 2004). This can be explained by the earlier snow melt due to a warmer climate. Is this warming of natural origin or is it a signal of global warming? We aim to answer that question in this study. If the change in hydrological variables is attributed to Greenhouse warming then we can compute how they will change in the future and that will allow planning and mitigation to proceed. It may also be possible to make formal Detection and Attribution (D&A) analysis statements (Hegerl, et al,; Allen and Tett) about other variables such as flood frequency, max/min temperatures, etc. that will be useful to policy makers.

A formal Detection and Attribution analysis requires information on the levels of natural variability expected in the detection variables, in this case stream flow and snow budget. These are generally obtained from a long control run of a global climate model (GCM). However, such control runs are made with relatively coarse resolution global models. Such models do not resolve topography well, especially something like the Sierra Nevada of California. Further, the stream flow simulations in these models are limited to the very largest rivers and the associated physics is crude. A final issue has to do with handling the enormous data volumes that would result from computations necessary to resolve the important physics of this problem. Few computational centers are equipped to handle the data volumes, realistically 100s of TB of data. For these reasons, D&A analysis cannot now be performed regionally at very high resolutions on the variables indicated above.

The Regional Climate analysis naturally falls into five components, described below:

1. Control run execution: The latest GCM from NCAR, the CCSM3, will be run in control run mode at LLNL on Thunder. The run will get us started on the problem of estimating natural variability for the eventual D&A analysis. The length of the control run will be 400 years, while its resolution will be approximately 1 degree latitude by 1 degree longitude grid (100km x 100km). If all the data from this run were saved it would amount to about 200TB/simulation year or 80,000 TB for the entire 400 year run. Clearly, we cannot handle that much data with existing resources, so some decisions will have to be made in collaboration with LLNL. The maintenance and execution of the large GCM run will be handled by LLNL. SIO will participate with LLNL in the setup of the run and in defining the scientific aspects of the run.

   SDSC will support the organization of the simulation output through use of data grid technology and accomplish the necessary data archiving. The current networks linking LLNL and UCSD are able to handle sustained data rates on the order of 300 Mbits/sec, implying the ability to move up to 3 TBs of data each day. If the run is done over 4 months of time, as much as 360 TBs of data could be moved. The amount archived will be limited by the availability of storage resources at SDSC. This proposal funds 100 TBs of archival storage space at SDSC. Possible ways to reduce the amount of data stored include limiting storage to selected spatial regions, and saving of periodic restart files. This requires the ability to pipeline the simulation data as it is generated through a set of data selection filters, organization of the selected data regions into a collection, and storage.

   Additional storage space may be needed at SDSC for the simulations. SDSC will collaborate with LLNL on the identification of alternate storage systems.

2. Downscaling: The 100km x 100km grid is still far too coarse to resolve the main features of interest. Accordingly, the global run will be downscaled to a 9 km grid using the MM5. This downscaling will be carried out in conjunction with the GCM run so that downscaled data will begin to become available almost immediately. Downscaling will be done over the western US…the entire US if feasible…. and also China, as serious climate change impacts are arising in this region also. The data volumes from the MM5 downscaling will be comparable to those from the global model run itself. Again, hard decisions and post processing will be needed to fit such data volumes into existing systems. SDSC will provide part of a person to work with LLNL/SIO on the storage and management issues and to accomplish the necessary data archiving. SDSC will provide data management tools to automate post processing and data management tasks. LLNL will be responsible for the execution of the MM5 runs. SIO will participate with LLNL in setting up the scientific requirement for the runs.

3. Hydrological modeling: The downscaled data on rainfall, temperature, terrain, etc. will be used to force a hydrological mode, most likely VIC, that gives time histories of steam flows and snow pack evolution (accumulation and melting) in the western US. The downscaled data for China will be reserved for future study. At a minimum, the resulting stream flow and other simulations will be checked against observations. Some statistical correction will likely be made to the key output variables. At the conclusion of this step, we will have all the information necessary to define the 'noise' or natural variability for the D&A analysis. SIO and LLNL will oversee and execute the VIC calculations. The simulation data will be managed and accessed using data grid technology provided by SDSC.
4. Steps 1-3 will be repeated with the GCM run in the forced mode. The forcing function definitions will be determined by SIO and LLNL working in conjunction with a panel of experts from NCAR, the same experts that constructed the model and who have run it in the forced mode for IPCC scenarios. With the anthropogenic forcing specified, at least 3 realizations of the GCM will be run for the period 1870-2100 to handle the ocean time lag. The alternative is to do a present day control run to serve as the starting point. A limited set of variables will be saved, essentially those needed to run the MM5 over the areas noted above and to study the hydrological cycle in these same areas. These calculations will be done by LLNL and use the data grid technology developed by SDSC. At the completion of this run we will have the information necessary to estimate the 'signal' or expected pattern of change in the hydrological variables projected to occur in a Greenhouse world.

5. The results of steps 3 and 4, taken together, provide us with all we need to undertake a full D&A analysis. This analysis will employ standard methodology and already exists at both SIO and LLNL. PIs at both locations will cooperate on this final phase of the project.

2.2 ENZO cosmology simulation data grid: The Cosmic Simulator:

Background: In the past decade, observational cosmology has become "big science" involving expensive instruments (e.g., the Hubble Space Telescope) and large teams (e.g., the Sloan Digital Sky Survey [SDSS]), attacking fundamental questions about the origin and evolution of the universe. Progress has been astonishing, including the discovery of the accelerating universe (Riess et al. 1998, Perlmutter et al. 1999), precision measurements of the global geometry, age, and composition of the universe (de Bernardis et al. 2000) and deep images of galaxies at the dawn of time (Beckwith et al. 2004). These and other observations have narrowed the range of acceptable theoretical models for cosmological structure formation to a single model called the *concordance model* (Bahcall et al. 1999) whose free parameters are now known to high precision (Spergel et al. 2003). Cosmology thus finds itself in a place not unlike particle physics, where the goal going forward is to refine and test the standard model with yet higher precision measurements. Fundamental science questions driving the field include the nature of dark energy and dark matter, the formation and evolution of galaxies and quasars, and how and when the intergalactic medium was re-ionized.

Future progress requires ambitious observational surveys of the universe of unprecedented depth and breadth. The SDSS is collecting megabytes of data per galaxy on nearly 1 million galaxies distributed throughout a volume of space many billions of light years on a side. Currently over 2 TB of data has been collected and archived, and this number is expected to grow to 5 TB by project's end. Several similarly sized surveys are underway, and much larger ones are planned. In particular, the Large aperture Synoptic Survey Telescope [LSST] will collect 15 TB of image data every night for a year, amassing a collection of tens of petabytes over several years. The LSST will produce an object catalog of a billion galaxies—a thousand fold increase over the SDSS. Coping with this “data flood” requires advanced scientific data management technologies.

In order to maximize the science return, results from massive surveys need to be compared to the detailed predictions of the concordance model. These take the form of massive cosmological simulations of the formation of galaxies and large scale structure. Just as Moore's Law is the force behind the data explosion in astronomy, it has also enabled numerical simulations of unprecedented size and complexity on massively parallel supercomputers.

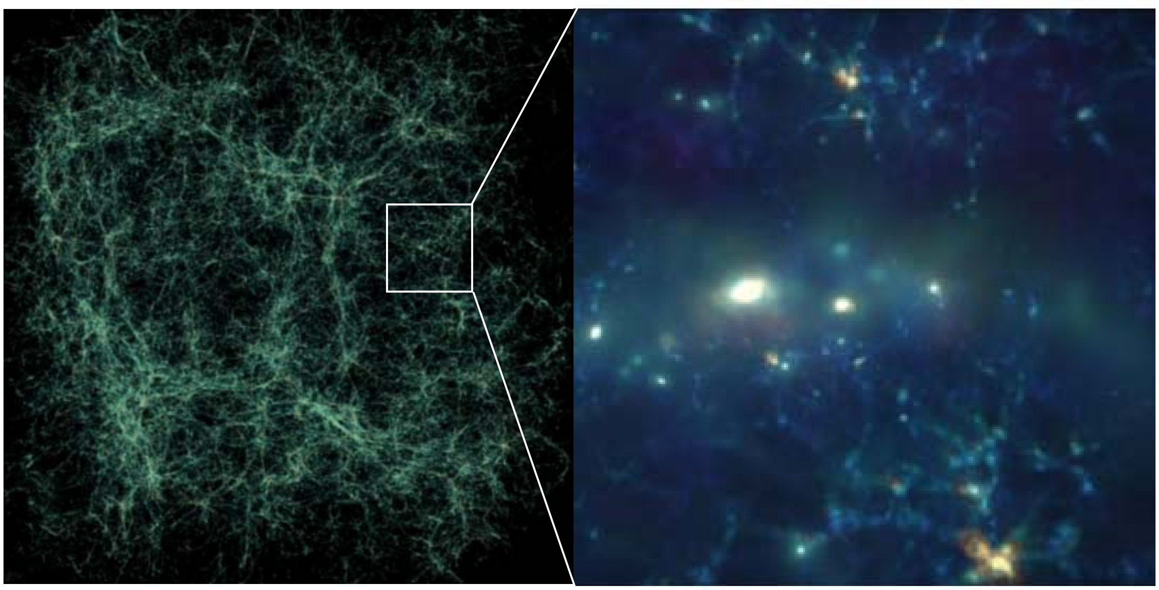

Figure 1. ENZO simulation of galaxy formation and large-scale structure. **Left**: survey volume simulated with 1 billion cells and particles on SDSC Blue Horizon. **Right**: Zoom in on young galaxies using adaptive mesh refinement (AMR).

ENZO is a parallel cosmology application developed at the Laboratory for Computational Astrophysics (LCA) at UCSD directed by Michael Norman. ENZO solves the equations of dark matter dynamics, multi-species hydrodynamics, non-equilibrium chemical and ionization kinetics, and self-gravity in an expanding universe dominated by dark energy. Parameterized models of star formation and feedback effects allow the simulation of the formation and evolution of galaxies on cosmic length scales and time scales. The state-of-the art is shown in Fig. 1. The simulation shown in the left panel evolves a concordance model with 1 billion Lagrangian dark matter particles and the equations of Eulerian hydrodynamics and self-gravity on a uniform grid of 1 billion

(1024^3) cells. The calculation was done on 512 processors of SDSC's IBM Blue Horizon computer, and produced 10TB of raw data and 6 TB of derived data. This calculation serves as a survey volume for follow-on adaptive mesh refinement (AMR) simulations which resolve the galaxies' internal structure. At right is shown an old AMR simulation of galaxy formation done at NCSA in 1998. Due to computer power and data handling limitations at the time, only 1/64 of the survey volume (256^3 base grid) could be simulated at high resolution. Now, with more powerful parallel computers and data management technologies, we can *in principle* simulate the entire volume at high spatial resolution. Making that a practical reality is the overarching goal of the cosmology simulation data grid project, which we shall henceforth refer to as the *Cosmic Simulator*.

<u>Project Goals:</u> The specific goals of the Cosmic Simulator project are threefold:
(1) use the LLNL-SDSC-UCSD data grid to be deployed to enable cosmological simulations of unprecedented size and physical realism;
(2) improve the physical realism of cosmological modeling through the inclusion of radiation transfer on adaptive meshes;
(3) generate simulated sky maps and galaxy catalogs using automated processing pipelines for LSST applications.

We elaborate on goals and means below.

<u>Use data grid to enable large AMR cosmological simulations</u>: We propose to use LLNL compute resources to carry out the largest AMR cosmological hydrodynamical simulation of galaxy formation and large scale structure ever attempted. The simulation will have a base grid resolution of 1024^3 cells and up to 8 levels of mesh refinement in galaxy forming regions. The total spatial dynamic range will be $1024 \times 2^8 \sim 250{,}000$. It is estimated the simulation will require 1 million cpu-hours on Thunder, and generate of order 100 TB of raw simulation output. Smaller scoping simulations will generate an order of magnitude less. The raw data will be archived and analyzed at SDSC through the proposed data grid. Using SDSC compute resources, the raw output will be converted into simulated images of the extragalactic sky and a simulated galaxy catalog.

Fig. 2 shows how simulated galaxy catalogs are produced from ENZO output. First, the observational survey parameters are translated into specifications for the ENZO initial conditions generator *Inits*, the output of which are read into ENZO for dynamical evolution. Periodic redshift dumps are the primary outputs and are archived in an SRB collection. Individual files, of order 10 GB each, containing baryonic fields, dark matter particles, and star particles are extracted for analysis and also archived. *Halo Finder* locates the dark matter halos which host the visible galaxies. *Profiler* computes integrated physical properties, which are output as an ASCII list. Given halo positions, the star particles are extracted as well. The star particles' attributes (mass, age, metallicity) are input to a spectral synthesis code which calculates each galaxy's luminosity, spectrum, and color. The spectral synthesis code takes as input a library of spectral energy distributions (SEDs) and telescopic filter functions, and outputs a list of the galaxies' observable properties. The physical and observable properties files are merged to form a simulated galaxy object catalog which will be archived with the raw and derived data in SRB. At present, these tasks are done manually as a post-processing step. The Cosmic

Simulator data grid will automate the execution of these steps via workflow tools described elsewhere.

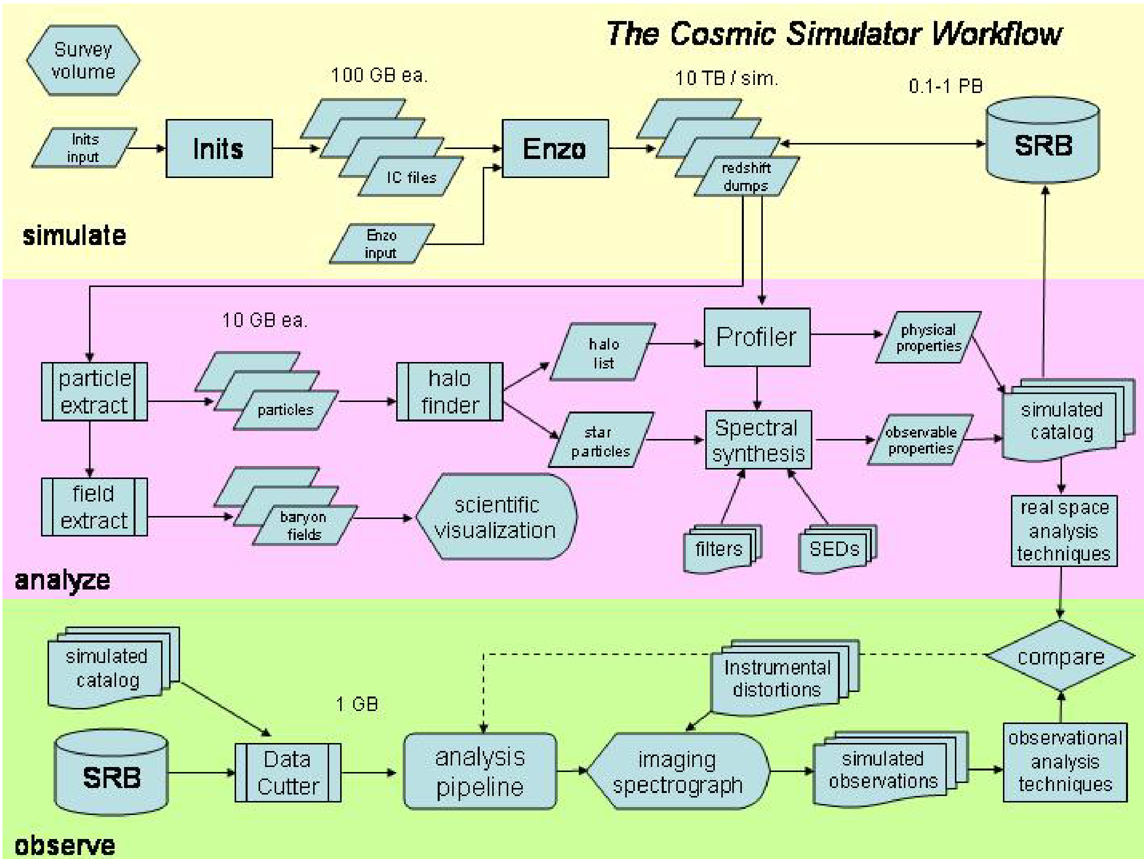

Figure 2. Cosmic Simulator workflow.

UCSD will work with LLNL computing personnel on porting and optimizing ENZO for large scale runs on Thunder and BlueGene/L. ENZO already runs well on the NSF TeraGrid clusters, which are architecturally similar to Thunder. SDSC has already installed a small BlueGene/L for code development, which Norman's group has access to.

Improve physical modeling: The physical modeling will be improved by adding the transport of ionizing radiation from young stars in galaxies. The latter is important for the process of self-regulated star formation, for predicting the epoch of cosmic re-ionization, and for predicting the observed properties of the earliest galaxies--the prime mission of the James Webb Space Telescope to be launched ca. 2010. LSST will complement the JWST with its enormous field of view. Norman's group will work with Frank Graziani at LLNL to verify and validate the AMR radiation transport module and radiation hydrodynamic tests. The development of appropriate linear system solvers for solving the transport equation on a structured adaptive mesh will be done in collaboration with Randy Bank (UCSD).

Simulate the LSST Sky: The LSST will image 10 square degrees of the sky at a time, reaching 24th magnitude in 10 seconds. Each 3 gigapixel image will contain millions of galaxies. By stacking these images much deeper exposures are obtained, increasing the galaxy count even further. By analyzing the weak lensing of the galaxies on the plane of the sky, one can deduce the distribution of dark matter in the universe on large scales. This is one of the principle science drivers of LSST. We will use ENZO to simulate a volume of the universe that is matched to LSST's 10 degree field of view at a redshift of 1. We will use the Cosmic Simulator pipeline to compute synthetic sky maps with a depth and resolution comparable to LSST. Using the data grid, the simulated sky maps will be provided to LLNL personnel involved in designing the LSST object analysis pipeline.

2.3 Scientific Data Management:

The ability to automate all aspects of data management has been a long-term research objective at SDSC. Data grid technology, called the Storage Resource Broker (SRB), has been developed that is used to enable data sharing across administrative domains, data publication through digital libraries, and data preservation through persistent archives. The Storage Resource Broker is middleware that manages interactions between a collection used to organize digital entities, and the underlying storage systems that hold the data sets and the databases that hold information about the data sets. The SRB maintains the consistency between provenance metadata, administrative metadata, authenticity metadata, and integrity metadata, and the data sets that are being organized. By controlling name spaces for resources, users, files, and metadata, the SRB is able to organize distributed data sets into a logical collection that is accessible over the Internet. The individual data sets can be distributed across multiple types of storage systems at multiple locations, and accessed through a preferred Application Programming Interface. This ability to manage a distributed collection makes it possible to integrate data resources at LLNL with data analysis environments at UCSD, while publishing data for use by a wider research community and preserving the data through replicas stored in archives.

The Storage Resource Broker is in production use at the San Diego Supercomputer Center. Over 140 Terabytes of simulation data from the ENZO cosmology code is currently archived at SDSC through the SRB. SIO earth systems observational data from ship voyages, remote sensors, and simulation code output are also stored in collections managed by the SRB. The total amount of data managed by the SRB at SDSC is over 350 Terabytes, comprising over 50 million files. The proposed collaboration will generate 100 TBs of data that will be archived at SDSC.

The advanced data management research and development tasks are focused on automating interactions between the storage systems, automating processing pipelines for generating derived data products and for detecting features, automating comparisons with observational data, and automating visualizations of both the simulation and observational data. The data management components needed to support automation include:

- Data grids to organize and manage the simulation output. The Earth System Grid at LLNL currently provides support for the climate simulation output. The SRB data grid at SDSC currently provides support for the cosmology simulation output.

- Data grid federation. The ability to replicate data between independent data grids has been demonstrated using the SRB. The feasibility of establishing a similar capability between the Earth System Grid and the SRB will be evaluated.
- Data processing pipelines. The ability to coordinate the execution of multiple processing steps is supported by the Chimera/Pegasus system, the Kepler/Ptolemy system, and the Matrix system. Chimera supports execution of CPU-intensive applications where data is moved to the computing platform. Kepler integrates functional actors for processing collections, and Matrix provides a dataflow language that can apply processing directly at remote storage systems. While several projects are examining the integration of all three systems to provide a comprehensive data processing pipeline, we will focus on the application of these systems to the ENZO and CCSM simulation outputs.
- Data visualization systems. The ability to support 3D rendering of massive (multi-terabyte sized) data sets has been demonstrated through the VISTA code at SDSC. VISTA has been used to generate visualizations of ENZO simulation output files.

SDSC will coordinate scientific data management development in support of the ENZO and CCSM simulations. The activities will include:

- integration of the SRB data grid with data caches at LLNL. SRB servers will be installed on selected storage repositories at LLNL, and on workstations used by researchers at UCSD and SIO.
- development of data movement capabilities for terascale applications. Bulk data movement and bulk metadata registration capabilities will be provided for moving data between LLNL and UCSD.
- integration with disk caches for enhanced analysis of simulation output. SDSC uses modular commodity disk storage systems (Grid Bricks) to cache data for analysis or for long-term publication. Such systems provide storage at a cost under $2000 per terabyte. Grid Bricks will be acquired and integrated into the analysis environments to facilitate data caching between LLNL and UCSD.
- integration with dataflow management system for managing interactions with data. SDSC will collaborate with LLNL on the implementation of data processing pipelines that are capable of analyzing collections of simulation output.
- integration with digital libraries for managing descriptions of features in the data. A digital library will be created for each scientific discipline (Astrophysics and Earth Systems Science) for publication of the simulation results and for correlation with observational data.
- integration with visualization systems for display of terascale 4D data sets.
- demonstration of prototype LSST data management pipeline. The scalability of the processing pipeline will be assessed for use within the Large Synoptic Survey Telescope project. An explicit example will be the generation of synthetic observational data to simulate what the LSST should see based upon the ENZO cosmology simulation results.
- development of terascale database applications. The number of features that may be detected within a simulation output data set can be orders of magnitude larger than the number of files. For instance, the number of stars and galaxies that will be detected by LSST will be on the order of 3 billion per observation. One approach is

to implement a catalog for each observation, and then develop analysis tools that work across multiple independent catalogs.

The scientific data management systems will be extended through collaborations with relevant disciplinary projects. These collaborations will be used to enable research projects that can grow into nationally-funded projects with lifetimes on the order of decades.

3. Project deliverables

From an overview perspective, for the global climate change application, we hope to determine if the substantial changes recently observed in key hydrological variables over the western US are the result of natural variability or due to anthropogenic forcing. For the cosmology simulation of the early universe we hope to simulate the observational data that would be seen by the LSST project. For the management of scientific data, we expect to demonstrate the management of scientific simulation collections that are hundreds of terabytes in size, and the organization of the material into digital libraries that support extended analysis.

The research activities that will be conducted during the three years of the grant include:

1st year

- Develop a standard scenario for climate modeling. We will define and complete a GCM control run. We will complete MM5 downscaling for selected geographic regions of the GCM control run.
- Begin comparison of MM5 data with observations.
- Develop forcing scenario for the GCM and a complete anthropogenic run.
- Begin VIC simulations with downscaled data.
- Implement a data grid linking resources between LLNL and SDSC. The data grid will be used to manage the simulation output that is generated.
- Port ENZO to LLNL compute resources and scaling studies
- Develop ENZO metadata schema and enhanced file I/O
- Develop Cosmic Simulator analysis pipeline and archive
- First simulated LSST sky map

2nd year

- Run and downscale the anthropogenic scenario.
- Complete VIC run on GCM/MM5 control run. Compare results with observations and develop statistical correctors as needed.
- Prepare paper on these results.
- Begin VIC runs with anthropogenically forced GCM/MM5 runs.
- Develop a digital library for publishing results, and possible integration with PCMDI
- Implement radiation transport in ENZO
- Execution of improved cosmology simulation code on LLNL resources
- Visualization of 4D output from the SDSC archive

3rd year

- Complete VIC runs with anthropogenic runs.
- Complete D&A analysis and a write a paper describing the results.
- Simulate improved LSST sky map
- Use Cosmic Simulator for weak lensing analysis
- Analysis of database scalability for large scientific collections

The simulation results generated by these applications will be stored at SDSC. The amount of results that can be stored will be determined by available funding for procuring storage systems. This proposal will support storage of 100 TBs at SDSC. Data transfers between LLNL and UCSD will be done through a LLNL link to the CENIC network. SDSC maintains a 10-Gbit/sec link within CENIC. The current LLNL link to CENIC through UC Berkeley is OC-12 (622 Mbits/sec). The amount of data moved between LLNL and UCSD will be limited to about 3 TBs per day. This will constrain the types of data analyses that can be done at UCSD to selected subsets of the simulation output.

4. Project Management:

The project will be coordinated by the three co-PIs:

- Reagan Moore will manage the team that develops and applies the data grid technology
- Michael Norman will manage the team that develops and executes the ENZO cosmology simulation.
- Tim Barnett will manage the team that develops and applies the global change simulations

The common activity between the three areas is the scientific data management required to support the simulation output. SDSC will provide a data coordinator to ensure that data is successfully organized into a collection and archived. In addition to actually archiving the data, this person will serve as a common knowledge source, ensuring smooth transitions across the various project interfaces that are data driven. Explicit project management activities include:

- SDSC/data coordinator. SDSC will work with LLNL personnel to interface to LLNL storage resources.
- SDSC/dataflow developer. SDSC will endeavor to automate the data management for both applications through integration of dataflow systems and data grid systems.
- LLNL to execute the global simulation and MM5 downscaling
- SIO to manage hydrological modeling.
- SIO/LLNL to jointly conduct D&A analysis.

5. Project sustainability

Collaborations will be pursued to continue the research beyond 3 years, to broaden participation to a national scale, and to broaden the application of the results. Examples include:

- Chronopolis preservation environment for NSF research data. A proposal has been submitted to the NSF to build a preservation facility for scientific data collections. One of the candidates for use of this facility is the LSST experiment. One goal is to demonstrate that LSST data collections can be economically housed in an NSF funded preservation facility.
- California Digital Library data depository for preservation of UC collections. This depository is intended to archive scientific research collections for the University of California system. One goal is to demonstrate the integration of publication environments for scientific data (digital libraries) between LLNL, the UC data depository, and the NSF data depositories.
- NSF National Virtual Observatory – integration of simulation and observational data. SDSC collaborates on the NVO project which is developing standard services for manipulating astrophysical data sets. Variants of the services developed within the NVO can be applied to general scientific data collections, including service registries, database access methods, and image access methods.
- LSST data management. SDSC is collaborating with the LSST project to demonstrate data management at scale. Data ingestion at nearly the design rate (10 TBs per day) has already been accomplished. A goal is to demonstrate that the technology used to support ENZO and Unified Climate Model output can also be used to support the LSST project.
- Cosmic Data Grid for cosmology community digital library (NSF). A proposal was submitted to NSF to build a national data grid to manage cosmology simulations. The proposed data grid is the national extension of the scientific data management system being developed in collaboration with LLNL.
- Climate-modeling Data Grid for analysis of climate change. The Earth Science Grid is based upon Globus technology, which differentiates services by functionality. The integration with the Storage Resource Broker data grid which differentiates services by scientific collection will broaden the reach of both data grids. A goal is to demonstrate that digital library technology can be built that supports access to data within both data grid environments.
- Teragrid scientific applications. The NSF supports execution of both climate and cosmology simulations on the Teragrid. At SDSC, compute platforms have been acquired for inclusion within the Teragrid that are compatible with the compute platforms at LLNL (Blue Gene/L, IBM p690 cluster). A goal is to demonstrate that the applications can run across both environments.
- Creation of community digital libraries for specific research topics. The publication of scientific results for use in both education and research is a goal of the NSF National Science Digital Library, for which SDSC operates a persistent archive. Simulation results for cosmology and global climate change are candidates for inclusion in the NSDL education repository.

This project opens several potential future avenues for research:

1. It opens the door to conducting further such studies on mammoth Earth Simulator class machines such as LLNL's Thunder and greatly expanded simulations on LLNL BGL. In the proposed program, we will be unique in conducting what will be the largest climate simulation yet done in the US. That should well position us for future

studies that might take a more global view of regional problems, e.g. regional interactions.

2. The western US results open the door to a broad range of collaborators concerned about water and energy management, etc. These will be the decision makers and planners who most need to know what the future holds.
3. Separate analyses of the China and India data sets, along the lines of those done in the western US, offer a broad basis for future funding and international collaboration.

6. Facilities Description:

The resources available at the San Diego Supercomputer Center include production data management systems as well as development environments for creating and testing the next generation software. The production data environment includes supercomputers, archival storage systems, high-performance SAN disk arrays, commodity-based disk systems (Grid Bricks), data management platforms, database platforms, and advanced visualization systems. The capabilities of the center include peak 10-teraflops-capable systems, a 9-petabyte archive, a 500-TByte SAN array, and high-performance data-handling systems that are capable of moving data at rates from 1 GB/sec (archival storage to SAN disk) to 5 GBs/sec (SAN disk to memory). SDSC is a node on the Teragrid, and has demonstrated movement of data between Teragrid nodes at rates of 500 MB/sec using parallel I/O streams managed by the Storage Resource Broker software.

The software systems include archival storage systems (IBM High Performance Storage System and SAM-QFS file system), Oracle and DB2 database technology, and the Storage Resource Broker data grid. The software systems are integrated into persistent archives for long term preservation, digital libraries for data publication, and data grids for data sharing in distributed environments.

The hardware and software systems are used in production to support the SIO-Explorer digital library, the Teragrid, the Real-time Observatories Network data grid, the National Science Digital Library persistent archive, the Joint Center for Structural Genomics data grid, the Alliance for Cell Signaling digital library, the UCSD library ArtStor image collection, the Southern California Earthquake Center digital library, and many others. The systems are also used to support testbeds for multiple NSF Information Technology Research projects, including the National Virtual Observatory, and the Geosciences Network. Systems at SDSC also support the National Archives and Records Administration research prototype persistent archive. In particular, the NPACI data grid provides a nation-wide data grid with 87 registered storage resources for the testing of distributed data management software at continental scales. The Storage Resource Broker data handling system uses a storage repository virtualization mechanism to ensure the ability to manage data stored in all types of storage repositories accessible to the NPACI data grid.

The facilities thus include not only hardware and software systems that can archive and manipulate the largest existing data collections, but also multiple production data collections that are in active use by earth science, bio-informatics, astronomy, education, and library communities. The provision of advanced visualization systems is important because this is one of the few ways that massive 10-TByte output files can be presented.

This requirement forced the development of 3-D visualization tools that are able to directly render images from files stored in the Storage Resource Broker data grid.

For software development, workstations are provided for senior staff. A 15-TByte Grid Brick commodity-based disk system is used to support collection development and software testing (at an effective cost of $2000 per TByte of disk). An Oracle database instance is used to support production systems. Non-proprietary database systems are used for testing, including PostgreSQL and mySQL. An additional commercial database, DB2, is used for production management of large catalogs.

The Teragrid production hardware and software systems are allocated under a national peer-review mechanism.

7. References:

1. Moore, R., "Persistent Collections," book chapter in "Databasing the Brain," editors S. H. Koslow and S. Subramaniam, John Wiley & Sons, 2005.
2. Moore, R., A. Rajasekar, M. Wan, "Data Grids, Digital Libraries and Persistent Archives: An Integrated Approach to Publishing, Sharing and Archiving Data", submitted to IEEE, Dec. 2004
3. Moore, R., "National Virtual Observatory Architecture," in "Toward an International Virtual Observatory: Proceedings of the ESO/ESA/NASA/NSF Conference Held at Garching Germany, 10-14 June 2002", editors Peter J. Quinn, Krzysztof M. Gorski.
4. Moore, R., "Digital Libraries and Data Intensive Computing," 2nd China Digital Library Conference, Beijing, China, September 2004.
5. Moore, R., "Integrating Data and Information Management," International Supercomputer Conference, Heidelberg, Germany, June 2004.
6. Rajasekar, A., M. Wan, R. Moore, W. Schroeder, "Data Grid Federation", PDPTA 2004 - Special Session on New Trends in Distributed Data Access, June 2004.
7. Jagatheesan, A., R., Moore, "Data Grid Management Systems," NASA / IEEE MSST2004, Twelfth NASA Goddard / Twenty-First IEEE Conference on Mass Storage Systems and Technologies, April 2004.
8. Moore, R., "Preservation Environments," NASA / IEEE MSST2004, Twelfth NASA Goddard / Twenty-First IEEE Conference on Mass Storage Systems and Technologies, April 2004.
9. Malaika, S., A. Merzky, R. Moore, "GGF Data Area Structure and Function Analysis", Global Grid Forum 10, Berlin, March 2004.
10. Moore, R., "Evolution of Data Grid Concepts", Global Grid Forum Data Area Workshop, January 2004.
11. Moore, R., "Operations for Access, Management, and Transport at Remote Sites," Global Grid Forum, December 2003.
12. Moore, R., A. Merzky, "Persistent Archive Concepts," Global Grid Forum, December 2003.
13. Moore, R., "Preservation of Data," e-Science Issue on Intelligent Systems, SDSC Technical Report 2003-006, San Diego, California, September, 2003.
14. Rajasekar, A., M. Wan, Reagan Moore, W. Schroeder, G. Kremenek, A. Jagatheesan, C. Cowart, B. Zhu, S.-Y. Chen, R. Olschanowsky, "Storage Resource

Broker - Managing Distributed Data in a Grid, "submitted to Computer Society of India Journal, special issue on SAN, 2003.
15. Moore, R., "Common Consistency Requirements for Data Grids, Digital Libraries, and Persistent Archives", Grid Protocol Architecture Research Group, Global Grid Forum, Tokyo, Japan, March 5, 2003.
16. Moore, R., C. Baru, "Virtualization Services for Data Grids", Book chapter in "Grid Computing: Making the Global Infrastructure a Reality", John Wiley & Sons Ltd, 2003.
17. Rajasekar, A., Michael Wan, Reagan Moore, George Kremenek, Tom Guptil, "Data Grids, Collections, and Grid Bricks", Proceedings of the 20th IEEE Symposium on Mass Storage Systems and Eleventh Goddard Conference on Mass Storage Systems and Technologies, San Diego, April 2003.
18. Wan, M., Arcot Rajasekar, Reagan Moore, Phil Andrews, "A Simple Mass Storage System for the SRB Data Grid", Proceedings of the 20th IEEE Symposium on Mass Storage Systems and Eleventh Goddard Conference on Mass Storage Systems and Technologies, San Diego, April 2003.
19. Rajasekar, A., Michael Wan, Reagan Moore, Arun Jagatheesan, George Kremenek, "Real Experiences with Data Grids - Case studies in using the SRB", International Symposium on High-Performance Computer Architecture, Kyushu, Japan, December, 2002.
20. Moore, R., "The San Diego Project: Persistent Objects," Archivi & Computer, Automazione E Beni Culturali, l'Archivio Storico Comunale di San Miniato, Pisa, Italy, February, 2003.
21. Berriman, G. Bruce, David Curkendall, John Good, Joseph Jacob, Daniel S. Katz, Mihseh Kong, Serge Monkewitz, Reagan Moore, Thomas Prince, Roy Williams, "An Architecture for Access to a Compute Intensive Image Mosaic Service in the NVO," SPIE Conference 4686 "Virtual Observatories", Hawaii, August 2002.
22. Rajasekar, A.,M. Wan, R. Moore, "mySRB and SRB, Components of a Data Grid", 11th High Performance Distributed Computing conference, Edinburgh, Scotland, July 2002.
23. Moore, R., "Preservation of Data, Information, and Knowledge," Proceedings of the World Library Summit, Singapore, April 2002.
24. Boisvert, R.,P. Tang, "The Architecture of Scientific Software," pp. 273- 284, "Data Management Systems for Scientific Applications," Kluwer Academic Publishers, 2001.
25. Chen, C., "Global Digital Library Development," pp. 197-204, "Knowledge-based Data Management for Digital Libraries," Tsinghua University Press, 2001.
26. Rajasekar, A.,R. Moore, "Data and Metadata Collections for Scientific Applications", High Performance Computing and Networking (HPCN 2001), Amsterdam, Holland, June 2001.
27. Stockinger, H.,O. Rana, R. Moore, A. Merzky, "Data Management for Grid Environments," European High Performance Computing and Networks Conference, Amsterdam, Holland, June 2001.

28. Moore, R., “Knowledge-based Grids,” Proceedings of the 18th IEEE Symposium on Mass Storage Systems and Ninth Goddard Conference on Mass Storage Systems and Technologies, San Diego, April 2001.
29. Ludäscher, B., R. Marciano, R. Moore, “Preservation of Digital Data with Self-Validating, Self-Instantiating Knowledge-Based Archives,” ACM SIGMOD Record, 30(3), 54-63, 2001.
30. Ludaescher, B.,R. Marciano, R. Moore, “Towards Self-Validating Knowledge-Based Archives,” SDSC TR 2001-1, January 2001.
31. Moore, R., “Knowledge-based Persistent Archives,” Proceedings of La Conservazione Dei Documenti Informatici Aspetti Organizzativi E Tecnici, in Rome, Italy, October 2000.
32. Brunner, R., S. Djorgovski, A. Szalay, “Virtual Observatories of the Future,” Astronomical Society of the Pacific Conference Series, Vol. 225, pp. 257-264, June 2000.
33. Moore, R., C. Baru, A. Rajasekar, B. Ludascher, R. Marciano, M. Wan, W. Schroeder, and A. Gupta, “Collection-Based Persistent Digital Archives – Parts 1& 2”, D-Lib Magazine, April/March 2000, http://www.dlib.org/
34. Foster, I., Kesselman, C., “The Grid: Blueprint for a New Computing Infrastructure,” Chapter 5, “Data Intensive Computing,” Morgan Kaufmann, San Francisco, 1999
35. Moore, R., J. Lopez, C. Lofton, W. Schroeder, G. Kremenek, M. Gleicher, “Configuring and Tuning Archival Storage System,” Proceedings of the 16th IEEE Symposium on Mass Storage Systems, March 1999.
36. Rajasekar, A., R. Marciano, R. Moore, “Collection Based Persistent Archives,” Proceedings of the 16th IEEE Symposium on Mass Storage Systems, March 1999.
37. Baru, C., R, Moore, A. Rajasekar, M. Wan,"The SDSC Storage Resource Broker,” Proc. CASCON'98 Conference, Nov.30-Dec.3, 1998, Toronto, Canada.
38. Riess, A. et al. 1998. “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant”, Astron. J., 116, 1009.
39. Perlmutter, S. et al. 1999. “Measurements of Omega and Lambda from 42 High-Redshift Supernovae”, Astrophys. J., 517, 565.
40. Beckwith, S. et al. 2003. “The Hubble Ultra Deep Field”, Bull. Amer. Astron. Soc., 202, 1705.
41. Spergel, D. et al. 2003. “First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters”, Astrophys. J. Supp., 148, 175.
42. Bahcall, N., Ostriker, J. P., Perlmutter, S. and Steinhardt, P. 1999. “The Cosmic Triangle: Revealing the State of the Universe”, Science, 284, 1481

8. Budget:

The co-PIs on the project are:

Reagan Moore (5% FTE) – coordinates data management research efforts. Moore is the director of the Data Intensive Computing Environment group at SDSC, and the PI on projects developing the Storage Resource Broker.
Tim Barnett (13% FTE) – coordinates research on global climate change.
Mike Norman (11% FTE) – coordinates research on cosmology Norman is director of Laboratory for Computational Astrophysics which developed ENZO.

The senior personnel involved in the data management research and development activities include:

Arun Jagatheesan (100% FTE) – development of data processing pipeline technology and coordination of simulation data management
Database (TBN) (93% FTE) – develop scalable database analysis techniques
Amit Chourasia (5% FTE) – apply parallel rendering algorithms for visualization of simulation output
Robert Harkness (50% FTE) – support ENZO application

The senior personnel involved in cosmology application include:

Pascal Paschos (Norman postdoc) (100% FTE) – simulate LSST sky; weak lensing analysis; radiative transfer algorithm development and testing
Dan Reynolds (Randy Banks postdoc) (100% FTE) – ENZO integration of AMR linear solvers; LLNL production runs
James Bordner (100% FTE) – cosmology simulation data analysis pipeline and archive; project website
Alexei Kritsuk (25%) – galaxy modeling

The senior personnel in the regional climate work will include:

Mike Dettinger (10%) hydrology
Dan Cayan (10%) hydrology
David Pierce (15%) computer applications, D&A analysis
Post Docs (2, 100%) TBN

Funding is included to support storage of 100 TBs of data on an SDSC archive, and to provide a 10 TB disk cache for on-line data access. The cost for each system is expressed in $/TB/year. The hardware systems and media are assumed to have either a 3-year lifetime or a 5-year lifetime. Costs are amortized for media, tape silo, tape drives, archive server, disks, and disk server. Prorated costs per TB are generated for software, hardware maintenance, electricity, and labor. Base on current assessments of technology, tape-based storage is available at a cost of $600/TB/year. The current fibre-channel attached disk storage costs $6500/TB/year. SATA disk-based storage will be available at a cost of $1400/TB/year later this year. Support for a database instance to manage the data grid costs $4000 per year. 100-TBs of tape storage, a 10-TB SATA disk cache, and a database instance will be dedicated to this project at a cost of $78,000 per year.

9. Resumes

Reagan Moore
Tim Barnett
Mike Norman

Reagan W. Moore

San Diego Supercomputer Center	Telephone: 858 534-5073
University of California, San Diego	Fax: 858 534-5152
9500 Gilman Drive, MC 0505	E-mail: moore@sdsc.edu
La Jolla, CA 92093-0505	

Professional Preparation

California Institute of Technology	BS Physics	1967
University of California, San Diego	MS Plasma Physics	1970
University of California, San Diego	Ph.D. Plasma Physics	1978

Appointments

2001	Distinguished Scientist, San Diego Supercomputer Center
1997	Adjunct Professor, CSE, University of California, San Diego
1994	Associate Director, Data Intensive Computing, San Diego Supercomputer Center
1988	Manager, Programming and Software Services, San Diego Supercomputer Center
1986	Manager, Systems Software, San Diego Supercomputer Center
1981	Staff Scientist, Plasma Theory Group, General Atomics
1976	Senior Research Scientist, Plasma Theory Group, General Atomics

Related Publications

1. R. Moore, A. Merzky, “Persistent Archive Concepts”, Global Grid Forum Persistent Archive Research Group, Global Grid Forum, Seattle, May, 2003.
2. R. Moore, C. Baru, “Virtualization Services for Data Grids”, Book chapter in "Grid Computing: Making the Global Infrastructure a Reality", John Wiley & Sons Ltd, 2003.
3. R. Moore, “The San Diego Project: Persistent Objects”, Proceedings of the Workshop on XML as a Preservation Language, Urbino, Italy, October 2002.
4. Moore, R., C. Baru, A. Rajasekar, B. Ludascher, R. Marciano, M. Wan, W. Schroeder, and A. Gupta, “Collection-Based Persistent Digital Archives – Parts 1& 2”, D-Lib Magazine, April/March 2000, http://www.dlib.org/
5. Foster, I., Kesselman, C., “The Grid: Blueprint for a New Computing Infrastructure,” Chapter 5, “Data Intensive Computing,” Morgan Kaufmann, San Francisco, 1999

Other Significant Publications

1. Rajasekar, A., M. Wan, R. Moore, W. Schroeder, “Data Grid Federation”, PDPTA 2004 - Special Session on New Trends in Distributed Data Access, June 2004.
2. Moore R., and A. Rajasekar, "Data and Metadata Collections for Scientific Applications", High Performance Computing and Networking (HPCN 2001), Amsterdam, Holland, June 2001.
3. Moore, R., “Knowledge-based Grids,” Proceedings of the 18th IEEE Symposium on Mass Storage Systems and Ninth Goddard Conference on Mass Storage Systems and Technologies, San Diego, April 2001.
4. Boisvert, R., P. Tang, “The Architecture of Scientific Software,” pp. 273- 284, “Data Management Systems for Scientific Applications,” Kluwer Academic Publishers, 2001.
5. Baru, C., R, Moore, A. Rajasekar, M. Wan, "The SDSC Storage Resource Broker,” Proc. CASCON'98 Conference, Nov.30-Dec.3, 1998, Toronto, Canada.

Synergistic Activities

- Service: NSF-OCE Ocean Information Technology Steering Committee (2003), IEEE Mass Storage System Technical Committee (1997-2003), NASA ESDISS advisory committee (2000-2003), NSF LIGO Project Advisory Committee (1997-1999), NASA Data Assimilation Office Advisory Committee (1997-1998).
- Leadership: Global Grid Forum Persistent Archive Research Group Chair (1999-2003), Chair of NSF Workshop on Knowledge Networking (1997), Co-chair SDSC Managing and Mining Massive Data Workshop (1995, 1998).
- Awards and Honors: U. S. Patent co-holder on "Elongated Toroid Fusion Device" (1988) and "Transparent management of data objects in containers" (2004), University of California Regents Fellowship (1967), member of Tau Beta Pi (1966)

Collaborators

Bruce Allen (Wisc. Milwaukee), James Allen (UMass), Martha Anderson (LC), Peter Arzberger (UCSD), Paul Avery (Florida Gainsville), Chaitanya Baru (SDSC), Paul Berkman (OSU), Francine Berman (UCSD), Julian Bunn (Caltech), Charlie Catlett (ANL), Anne Chervenak (USC), Ewa Deelman (USC), Thomas DeFanti (UIC), Mark Ellisman (UCSD), Ian Foster (U. Chicago/ANL), Dave Fulker (UCAR), Hector Garcia-Molina (Stanford), John Good (IPAC/Caltech), Daniel Greenstein (UCOP-CDL), Susan Graham (UCB), Amarnath Gupta (UCSD), Bob Hanisch (JHU), Walter Hoehn (Columbia), Robert Hollebeek (Penn), Greg Janee (UCSB), Joseph JaJa (Maryland), William Johnston (LBNL), Tom Jordan (USC), Sidney Karin (UCSD), Carl Kesselman (USC), Dean Krafft (Cornell), Albert Lazzarini (Caltech), Carl Lagoze (Cornell), Miron Livny (Wisc. Madison), Jane Mandelbaum (LC), Richard Marciano (SDSC), William Martin (Michigan), Keith Marzullo (UCSD), Paul Messina, Bernard Minster (USCD/SIO), Richard Mount (SLAC), Harvey Newman (Caltech), J. Tinsley Oden (Texas, Austin), Arthur Olson (SRI), Andreas Paepcke (Stanford), Cherri Pancake (OSU), Yannis Papakonstantinou (UCSD), Wayne Pfeiffer (UCSD), Ruth Pourdes (FNAL), Lawrence Price (ANL), Tom Prince (Caltech), Mark Pullen (GMU), Sanguthevar Rajasckaran (Florida Gainsville), Arcot Rajasekar (SDSC), Dan Reed (UIC), Joseph Romano (Texas Brownsville), Arie Shoshani (LBNL), Terence Smith (UCSB), Rick Stevens (ANL), Shankar Subramanian (UCSD), Alexander Szalay (Johns Hopkins), Peter Taylor (University of Warwick), Kenneth Thibodeau (NARA), Doug Tody (NOAO), Robert Wilensky (UCB), Roy Williams (Caltech), Rich Wolski (UCSB).

GRADUATE STUDENTS AND POSTDOCTORAL RESEARCHERS

Xi Cynthia Sheng (UCSD)
Qiao Xin (UCSD)
Marcio Faerman (UCSD)
Ullas Kapadia (U Texas)
Jonathan Weinburg (UCSD)
Yi Li (UCSD)
Yuanfang Hu (UCSD)

GRADUATE AND POSTGRADUATE ADVISORS

William Thompson, UC San Diego

TIM P. BARNETT

EDUCATION: B.A., Physics, Pomona College
M.S., Physical Oceanography,
Scripps Institution of Oceanography,
Ph.D., Physical Oceanography,
Scripps Institution of Oceanography

PROFESSIONAL EXPERIENCE: July 1971 - present
Research Marine Physicist
Academic Administrator (1971-1982)
Scripps Institution of Oceanography
University of California, San Diego
La Jolla, California, 92093-0224

SELECTED HONORS AND AWARDS

- Elected to International Association for the Physical Sciences of the Ocean (IAPSO) Committee on Tides and Mean Sea Level for three-year term (1983-1986).
- Member NAS Committee on Climate Change and Joint Organizing Committee, WCRP
- Elected, Councilor, 3 year term, American Meteorological Society, 1/92-1/95
- Elected Fellow, American Meteorological Society, Fall, 1991
- Special Creativity Award, National Science Foundation, 1991 and 1992
- Awarded Sverdrup Gold Medal, (AMS), Fall, 1992
- Invited Congressional Testimony, "Preparing for El Nino", Sept. 11, 1997
- Invited Speaker, California State Legislature, "El Niño '97-'98: Prediction of Impacts in California", Sacramento, Jan. 22, 1998.
- San Diego Press Club Headliner of the Year 1998 (for El Nino work)
- Elected Fellow, American Geophysical Union, January, 2001

SELECTED OTHER PROFESSIONAL ACTIVITIES

- Numerous interviews with the press and TV nationally and internationally regarding weather, climate, El Nino, and greenhouse effects.
- Maintain scientific exchange program with Max Planck Institut, Hamburg, W. Germany
- Advise different branches of government (e.g., NOAA, DOE, NASA, JPL, EPA) on climate research and remote sensing.
- Over 200 talks to scientific and lay groups

SELECTED PUBLICATIONS

Barnett, T.P., E. Kirk, M. Latif, and E. Roeckner, 1991: On ENSO physics, *J. Clim.*, **4,** 487-515.

Barnett, T.P., L. Bengtsson, K. Arpe, M. Flugel, N. Graham, M. Latif, J. Ritchie, E. Roeckner, U. Schlese, U. Schulzweida, and M. Tyree, 1994: Forecasting global ENSO-related climate anomalies, *Tellus*, **46A**, 381-397. .

Barnett, T.P., 1995: Monte Carlo climate forecasts. *J. Clim.,* **8(5)**, 1005-1022.

Latif, M. and T.P. Barnett, 1996: Decadal climate variability over the North Pacific and North America: Dynamics and predictability. *J. Clim.,* **9,** 2407-2423.

Barnett, T.P., G. Hegerl, B. Santer and K. Taylor, 1998: The potential effect of GCM uncertainties on greenhouse signal detection, *J. Clim.*, **11(4)**, 659-675. Barnett, T.P., K. Hasselmann, M. Chelliah, T. Delworth, G. Hegerl, P. Jones, E. Rasmusson, E. Roeckner, C. Ropelewski, B. Santer, and S. Tett, 1999: Detection and attribution of recent climate change, *Bull Am. Met. Soc.*, **80(12)**, 2631-2659.

Barnett, T.P., R. Schnur, and D.W. Pierce, 2001: Detection of anthropogenic climate change in the world's oceans, *Science*, **292**, 270-274.

MICHAEL L. NORMAN

Preparation: California Institute of Technology, Astronomy, B.S. 1975
UC Davis, Engineering & Applied Science, M.S. 1976
UC Davis, Engineering & Applied Science, Ph.D. 1980
Max-Planck-Institut fur Astrophysik, Postdoc, 1980-1984

Appointments: Professor of Physics, UC San Diego, 2000{present
Senior Fellow, San Diego Supercomputer Center, 2000{present
Adjunct Professor of Astronomy, U. Illinois, Urbana, 2000-present
Alexander von Humboldt Research Award Scholar, FRG 1997-1998
Professor of Astronomy, U. Illinois, Urbana, 1991-2000
Senior Research Scientist, NCSA, U. Illinois, Urbana, 1989-2000
Associate Director, NCSA, U. Illinois, Urbana, 1987-1989
Research Scientist, NCSA, U. Illinois, Urbana, 1986-1987
Staff Member, Los Alamos Nat'l Lab, 1984{1986

Synergistic Activities: (1) Norman directs the Laboratory for Computational Astrophysics which develops and disseminates application codes (ZEUS-2D, ZEUS-3D, ZEUS-MP, KRONOS, ENZO) for astrophysics research. (2) Norman has also been involved in many public outreach projects through videos and lms featuring astrophysical simulations, including IMAX lm Cosmic Voyage, two NYC Hayden Planetarium Sky Shows, Nova Runaway Universe, and Discovery Channel Unfolding Universe

Advisors: PhD: James R. Wilson (LLNL), postdoc: Karl-Heinz Winkler (LANL)

PhD students: David Clarke (St. Marys), James Stone (Princeton), Dinshaw Balsara (Notre Dame), Byung-Il Jun (LLNL), Yu Zhang (Prowess Systems), Greg Bryan (Oxford), Henry Neeman (U. Oklahoma), Tom Abel (Penn State), Wen-Ching Lin (UCSD), Tridivesh Jena (UCSD)

Postdocs: David Clarke (St. Marys), James Stone (Princeton), R. Fiedler (UIUC), Doug Swesty (SUNY), John Hayes (UCSD), Alexei Razoumov (ORNL), James Bordner (UCSD), Greg Daues (NCSA), Pakshing Li (UCB)

Collaborators: T. Abel (Penn State), J. Burns (U. Colorado), G. Bryan (Oxford), R. Cen (Princeton), R. Dav e (Princeton), C. Frenk (Durham), L. Hernquist (Harvard), A. Kritsuk (UCSD), C. Loken (CITA), P. Madau (UCSC), A. Meiksin (Edinburgh), J. Ostriker (Princeton), D. Weinberg (OSU), S. White (MPA), M. White (UCB)

Five publications relevant to the proposed research

1. Norman, M.L. 2004. \The Impact of AMR in Numerical Astrophysics and Cosmology", in Adaptive Mesh Re nement - Theory and Applications, Eds. T. Plewa, T. Linde & V. G. Weirs, Springer Lecture Notes in Computational Science and

Engineering.
2. O'Shea, B.W., Bryan, G.L., Bordner, J., Norman, M.L., Abel, T., Harkness, R., & Kritsuk, A. 2004. \Introducing Enzo, and AMR Cosmology Application", in Adaptive Mesh Refinement - Theory and Applications, Eds. T. Plewa, T. Linde & V. G. Weirs, Springer Lecture Notes in Computational Science and Engineering.
3. O'Shea, B.W., Nagamine, K., Springel, V., Hernquist, L., & Norman, M.L. 2004. “Comparing AMR and SPH Cosmological Simulations: I. Dark Matter & Adiabatic Simulations", Ap. J., in press. (astro-ph/0312651)
4. Tassis, K., Abel, T., Bryan, G.L., and Norman, M.L. 2003 \Numerical Simulations of High Redshift Star Formation in Dwarf Galaxies", ApJ, 587, 13.
5. Norman, M. L. & Bryan, G. L. 1999. \Cosmological Adaptive Mesh Refinement", in Numerical Astrophysics 1998, eds. S. Miyama & K. Tomisaka, (Kluwer, Dordrecht), 17.

Five additional publications

1. Norman, M.L. 2003. \Cosmological Simulations of X-ray Clusters: The Quest for Higher Resolution and Essential Physics", in Matter and Energy in Clusters of Galaxies, Eds. S. Bowyer & C.-Y. Hwuang, PASP Conference Series Vol. 301
2. Razoumov, A.O., Norman, M.L., Abel, T. and Scott, D. 2002, \Cosmological Hydrogen Reionization with Three-Dimensional Radiative Transfer", ApJ, 572, 695{704
3. Abel, T., Bryan, G.L. and Norman, M.L. 2002, \The Formation of the First Star in the Universe", Science, 295, 93{98
4. Norman, M.L., Daues, G., Nelson, E., Loken, C., Burns, J., Bryan, G.L. & Klypin, A. 2000. \ Simulated Cluster Archive: A Computational Catalog of X-Ray Clusters in a Lambda-CDM Universe", in Large-Scale Structure in the X-Ray Universe, Eds. M. Plionis & I. Georgantopoulos, (Atlantisciences, Athens).
5. Abel, T., Norman, M. L. & Madau, P. 1999. \Photon Conserving Radiative Transfer around Point Sources in Multi-Dimensional Numerical Cosmology", ApJ, 523, 66.