Metacomputing

Larry Smarr
Charles E. Catlett

From the standpoint of the average user, today's computer networks are extremely primitive compared to other networks. While the national power, transportation, and telecommunications networks have evolved to their present state of sophistication and ease of use, computer networks are at an early stage in their evolutionary process. Eventually, users will be unaware they are using any computer but the one on their desk, because it will have the capability to reach out across the national network and obtain whatever computational resources are necessary.

The computing resources transparently available to the user via this networked environment have been called a *metacomputer*. The metacomputer is a network of heterogeneous, computational resources linked by software in such a way that they can be used as easily as a personal computer. In fact, the PC can be thought of as a minimetacomputer, with a general-purpose microprocessor, perhaps floating point-intensive coprocessor, a computer to manage the I/O—or memory—hierarchy, and a specialized audio or graphics chip. Like the metacomputer, the minimetacomputer is a heterogeneous environment of computing engines connected by communications links. Driving the software development and system integration of the NCSA metacomputer are a set of "probe" metaapplications.

The first stage in constructing a metacomputer is to create and harness the software to make the user's job of utilizing different computational elements easier. For any one project, a typical user might use a desktop workstation, a remote supercomputer, a mainframe supporting the mass storage archive, and a specialized graphics computer. Some users have worked in this environment for the past decade, using *adhoc*, custom solutions, providing specific capabilities at best, in most cases moving data and porting applications by hand from machine to machine. The goal of building a metacomputer is elimination of the drudgery involved in carrying out a project on such a diverse collection of computer systems. This first stage is largely a software and hardware integration effort. It involves interconnecting all of the resources with high-performance networks, implementing a distributed file system, coordinating user access across the

various computational elements, and making the environment seamless using existing technology. This stage is well underway at a number of federal agency supercomputer centers.

The next stage in metacomputer development moves beyond the software integration of a heterogeneous network of computers. The second phase involves spreading a single application across several computers, allowing a center's heterogeneous collection of computers to work in concert on a single problem. This enables users to attempt types of computing that are virtually impossible without the metacomputer. Software that allows this to be done in a general way (as opposed to one-time, *ad hoc* solutions) is just now emerging and is in the process of being evaluated and improved as users begin to work with it.

The evolution of metacomputing capabilities is constrained not only by software but also by the network infrastructure. At any one point in time, the capabilities available on the local area metacomputer are roughly 12 months ahead of those available on a wide-area basis. In general, this is a result of the difference between the network capacity of a local area network (LAN) and that of a wide-area network (WAN). While the individual capabilities change over time, this flow of capabilities from LAN to WAN remains constant.

The third stage in metacomputer evolution will be a transparent national network that will dramatically increase the computational and information resources available to an application. This stage involves more than having the local metacomputer use remote resources (i.e., changing the distances between the components). Stage three involves putting into place both adequate WAN infrastructure and developing standards at the administrative, file system, security, accounting, and other levels to allow multiple LAN metacomputers to cooperate. While this third epoch represents the five-year horizon, an early step toward this goal is the collaboration between the four National Science Foundation (NSF) supercomputer centers to create a "national virtual machine room." Ultimately, this will grow to a truly national effort by encompassing any of the attached National Research and Education Network (NREN) systems. System software must evolve to transparently handle the identification of these resources and the distribution of work.

In this article, we will look at the three stages of metacomputing, beginning with the local area metacomputer at the National Center for Supercomputing Applications (NCSA) as an example of the first stage. The capabilities to be demonstrated in the SIGGRAPH '92 Showcase environment represent the beginnings of the second stage in metacomputing. This involves advanced user interfaces that allow for participatory computing as well as examples of capabilities that would not be possible without the underlying stage-one metacomputer. The third phase, a national metacomputer, is on the horizon as these new capabilities are expanded from the local metacomputer out onto Gbit/sec network testbeds.

LAN Metacomputer at NCSA

Following the PC analogy, the hardware of the LAN metacomputer at NCSA consists of subcomponents to handle processing, data storage and management, and user interface, with high-performance networks to allow communication between subcomponents (see Figure 1). Unlike the PC, the subsystems now are not chips or dedicated controllers, but entire computer systems whose software has been optimized for its task and communication with the other components. The processing unit of the metacomputer is a collection of systems representing today's three major architecture types: massively parallel (Thinking Machines CM-2 and CM-5), vector multiprocessor (CRAY-2, CRAY Y-MP, and Convex systems), and superscalar (IBM RS/6000 systems and Silicon Graphics (SGI) VGX multiprocessors). More generally, these are differentiated as shared memory. (Crays, Convex, and SGI) and distributed memory (CM-2, CM-5, and RS/6000s) systems.

Essential to the Phase I LAN metacomputer is the development of new software allowing the program applications planner to divide applications into a number of components that can be executed separately, often in parallel, on a collection of computers. This requires both a set of primitive utilities to allow low-level communications between parts of the code or *processes*, and the construction of a programming environment that takes available metacomputer resources into account during the design, coding, and execution phases of an application's development. One of the problems faced by the low-level communications software is that of converting data from one system's representation to that of a second system. NCSA has approached this problem through the creation of the Data Transfer Mechanism (DTM), which provides message-based interprocess communication and automatic data conversion to applications programmers and to designers of higher-level software development tools.[1]

At the level, above interprocess communication, there is a need for standard packages that help the applications designer parallelize code, decompose code into functional units, and spread that distributed application onto the metacomputer. NCSA's approach to designing a distributed applications environment has been to acquire and evaluate several leading packages for this purpose, including Parallel Virtual Machine (PVM),[2] and *Express*,[3] both of which allow the programmer to identify subprocesses or subsections of a dataset within the application and manage their distribution across a number of processors, either on the same

physical system or across a number of networked computational nodes. Other software systems NCSA is investigating include Distributed Network Queueing System (DNOS)[4] and Network Linda.[5] The goal of these efforts is to prototype distributed applications environments, which users can either use on their own LAN systems or use to attach NCSA computational resources when appropriate. Demonstrations in Showcase will include systems developed in these environments.

A balanced system is essential to the success of the metacomputer. The network must provide connectivity at application-required bandwiths between computational nodes, information and data storage locations, and user interface resources, in a manner independent of geographical location.

The national metacomputer, being developed on Gbit network testbeds such as the BLANCA testbed illustrated in Figure 2, will change the nature of the scientific process itself by providing the capability to collaborate with geographically dispersed researchers on Grand Challenge problems. Through heterogeneous networking technology, interactive communication in real time—from one-on-one dialogue to multiuser conferences—will be possible from the desktop. When the Internet begins to support capacities at 150Mbit/sec and above, commensurate with local area and campus area 100Mbit/sec FDDI networks, then remote services and distributed services will operate at roughly the same level as local services. This will result in the ability to extend local-area metacomputers to the national scale.

Metacomputing at SIGGRAPH '92 Showcase

The following descriptions represent a cross-section of a variety of capabilities to be demonstrated by metaapplication developers from many different institutions. These six applications also cut across three fundamental areas of computational science. *Theoretical simulation* can be thought of as using the metacomputer to solve scientific equations numerically. *Instrument/Sensor control* can be thought of as using the metacomputer to translate raw data from scientific instruments and sensors into visual images, allowing the user to interact with the instrument or sensor in real time as well. Finally, *Data Navigation* can be thought of as using the metacomputer to explore large databases, translating numerical data into human sensory input.

Theoretical Simulation

Theoretical simulation is the use of high-performance computing to perform numerical experiments, using scientific equations to create an artificial numerical world in the metacomputer memory where experiments take place without the constraints of space or time. One of these applications takes advantage of emerging virtual reality technologies to explore molecular structure, while a second theoretical simulation application we described allows the user to explore the formation and dynamics of severe weather systems. An important capability these applications require of the metacomputer is to easily interconnect several computers to work on a single problem at the same time.

Molecular Virtual Reality: This project will demonstrate the interaction between a virtual reality system and a molecular dynamics program running on a Connection Machine. Molecular dynamics models, developed by Klaus Schulten and his colleagues at the University of Illinois at Urbana-Champaign's Beckman Institute Center for Concurrent Biological Computing, are capable of simulating the ultrafast motion of macromolecular assemblies such as proteins.[6] The new generation of parallel machines allows one to rapidly simulate the response of biological macromolecules to small structural perturbations, administered through the virtual reality system, even for molecules of several thousand atoms.

Schulten's group, in collaboration with NCSA staff, developed a graphics program which collects the output of a separate program running on a Connection Machine and renders it on a Silicon Graphics workstation. The imagery can be displayed on the Fake Space Labs *boom* display system, VPL's EyePhone head-mounted display, or the Silicon Graphics workstation screen. The program provides the ability to interact with the molecule using a VPL DataGlove. The DataGlove communicates alterations of the molecular structure to the Connection Machine, restarting the dynamics program with altered molecular configurations.

This metaapplication will provide the opportunity to use Virtual Reality (VR) technology to monitor and control a simulation run on a Connection Machine stationed on the show floor. In the past, remote process control has involved starting, stopping, and changing the parameters of a numerical simulation. The VR user interface, on the

[1]DTM was developed by Jeff Terstriep at NCSA as part of the BLANCA Testbed efforts. NCSA's research on the BLANCA testbed is supported by funding from DARPA and NSF through the Corporation for National Research Initiatives.

[2]PVM was developed by a team at Oak Ridge National Laboratory, University of Tennessee, and Emory University. Also see A. Beguelin, J. Dongarra, G. Geist, R. Manchek, and V. Sunderam. Solving Computational Grand Challenges Using a Network of Supercomputers. In *Proceedings of the Fifth SIAM Conference on Parallel Processing*, Danny Sorenson, Ed., SIAM, Philadelphia, 1991.

[3]Express was developed at CalTech and is now being distributed by ParaSoft. It is a suite of tools similar to PVM.

[4]"DNQS, A Distributed Network Queueing System," and "DQS, A Distributed Queueing System," are both 1991 papers by Thomas Green and Jeff Snyder from SCRI/FSU. DNQS was developed at Florida State University.

[5]Network Linda was developed at Yale University.

[6]This research is by Mike Krogh, Rick Kufrin, William Humphrey and Klaus Schulten Department of Physics, National Center for Supercomputing Applications at Beckman Institute.

Figure 1.

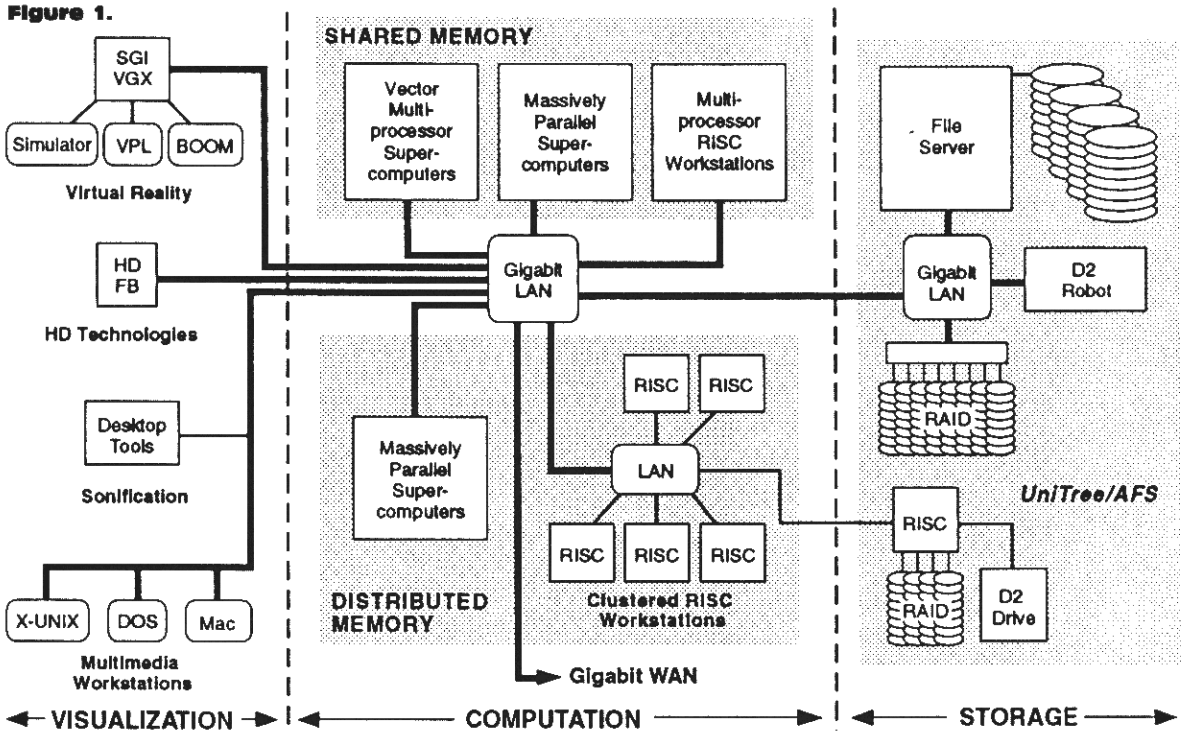


Figure 2.

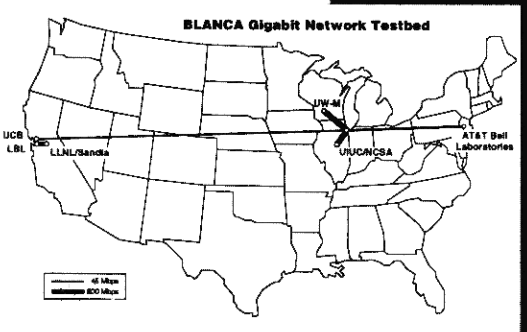


Figure 3.

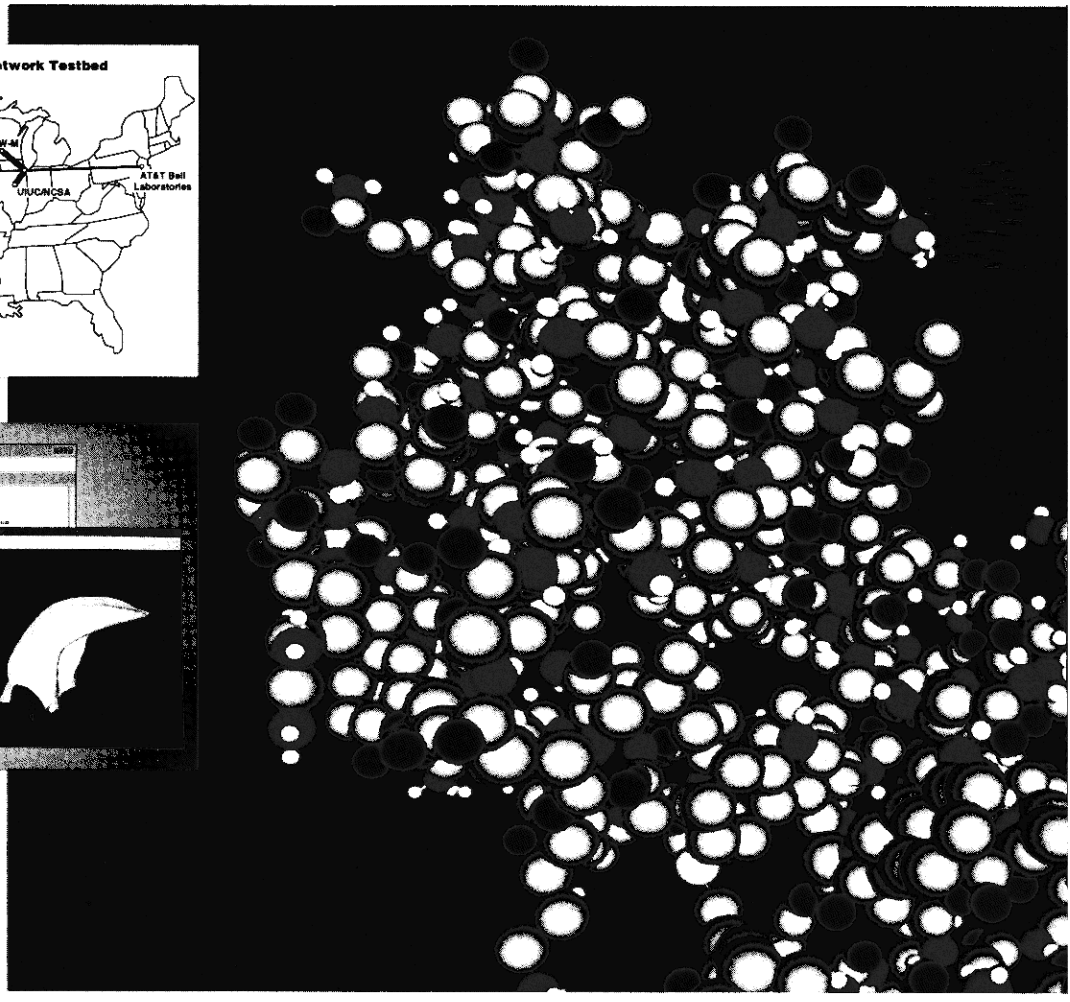
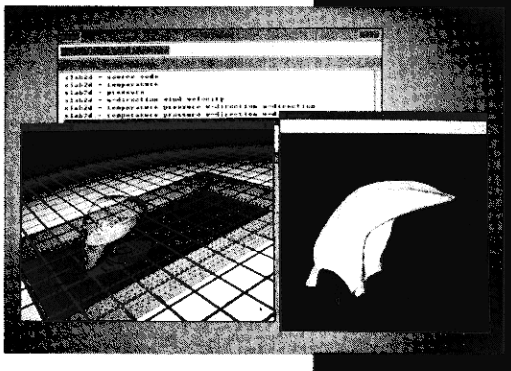


Figure 4.



other hand, allows the user to interact with and control the objects within the model—the molecules themselves—rather than just the computer running the model.

User-Executed Simulation/Analysis of Severe Thunderstorm Phenomena: In an effort to improve weather prediction, atmospheric science researchers are striving to better understand severe weather features. Coupled with special observing programs are intense numerical modeling studies that are being used to explore the relationship between these features and larger-scale weather conditions.[7] A supercomputer at NCSA will be used to run the model, and several workstations at both NCSA and Showcase will be used to perform distributed visualization processing and user control.

In Showcase, the visitor will be able to explore downburst evolution near the ground through coupled model initiation, simulation, analysis, and display modules. In this integrated, real-time environment, the analysis modules and visual display will be tied to new flow data as it becomes available from the model. This is a precursor to the kind of metacomputer forecasting environment that will couple observations, model simulations, and visualization together. The metacomputer is integral to the future forecasting environment for handling the large volumes of data from a variety of observational platforms and models being used to 'beat the real weather'. In the future, it is possible that real-time Doppler data will be used to initialize storm models to help predict the formation of tornadoes 20 to 30 minutes ahead of their actual occurrence.

Figure 1. LAN metacomputer at NCSA

Figure 2. BLANCA research participants include the University of California—Berkeley, Lawrence Livermore National Laboratories, University of Wisconsin-Madison (CS, Physics, Space Science and Engineering Center), and University of Illinois at Urbana-Champaign (CS, NCSA). Additional XUNET participants include Lawrence Livermore National Laboratories and Sandia. BLANCA uses facilities provided by the AT&T Bell Laboratories XUNET Communications Research Program in cooperation with Ameritech, Bell Atlantic, and Pacific Bell. Research on the BLANCA testbed is supported by the Corporation for National Research Initiatives with funding from industry, NSF, and DARPA. Diagram: Charles Catlett.

Figure 3. Three-dimensional image of molecule modeled with molecular dynamics software.
Credit: Klaus Schulten, NCSA visualization group

Figure 4. Comparing video images (background) with live three-dimensional output from thunderstorm model using the NCSA digital library.
Credit: Bob Wilhelmson, Jeff Terstriep

Instrument/sensor Control

Whereas the numerical simulation data came from a computational model, the data in the following applications comes from a scientific instrument. Now that most laboratory and medical instruments are being built with computers as control devices, remote observation and instrument control is possible using networks.

Interactive Imaging of Atomic Surfaces: The scanning tunneling microscope has revolutionized surface science by enabling the direct visualization of surface topography and electronic structure with atomic spatial resolution. This project will demonstrate interactive visualization and distributed control of remote imaging instrumentation.[8] Steering imaging experiments in real time is crucial, as it enables the scientist to optimally utilize the instrument for data collection by adjusting observation parameters during the experiment. A scanning tunneling microscope (STM) located in the Beckman Institute for Advanced Science and Technology at UIUC will be remotely controlled from a workstation at Showcase '92. The STM data will be sent as it is acquired to a Convex C3800 at NCSA for image-processing and visualization. This process will occur during data acquisition. STM instrument and visualization parameters will be under user control from a workstation at Showcase. The user will be able to remotely steer the STM in Urbana from Chicago and visualize surfaces at the atomic level in real time.

The project will use AVS, from AVS, Inc. to distributed components of the application between the Convex C3800 at NCSA and a showcase workstation. Viewit, a multidimensional visualization interface, will be used as the user interface for instrument control and imaging.

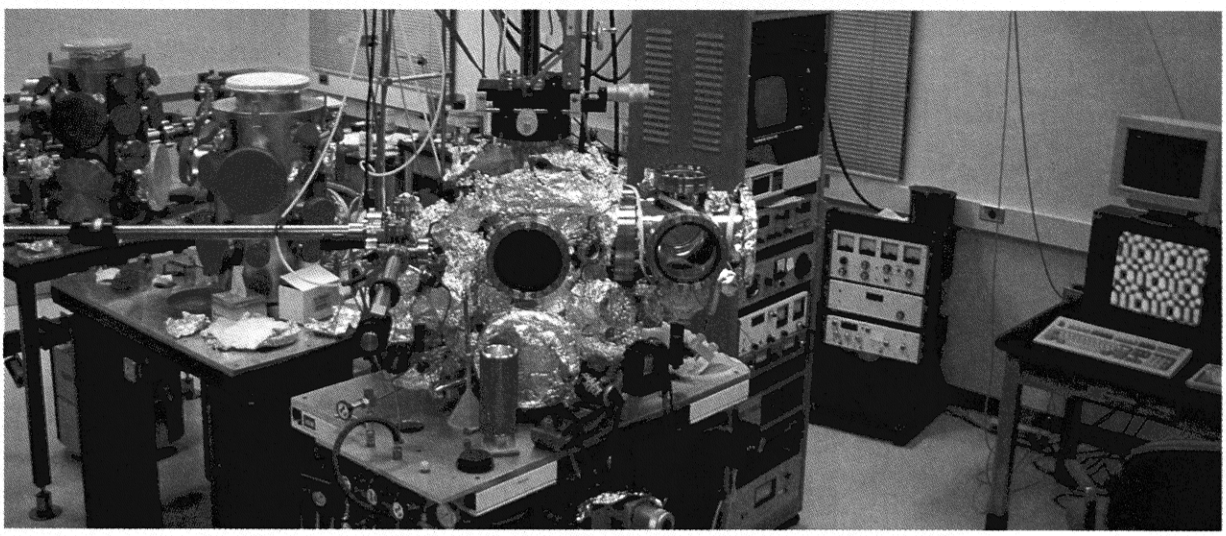
Data Navigation

Data navigation may be regarded not only as a field of computational science but as the method by which all computational science will soon be carried out. Both theoretical simulation and instrument/sensor control produce large sets of data that rapidly accumulate over time. Over the next several years, we will see an unprecedented growth in the amount of data that is stored as a result of theoretical simulation, instruments and sensors, and also textual and image data through network-based publication and collaboration. While the three previous applications involve user interfaces to specific types of data, the three following applications address the problem faced by scientists who are searching through many types of data. Capabilities are shown for solving the problem of locating data as well as examining the data.

Interactive Four-Dimensional Imaging: There are many different methods for visualizing biomedical image data sets. For instance, the Mayo Clinic Dynamic Spatial Reconstructor (DSR) is a CT scanner which can collect entire three-dimensional scans of a subject as quickly as 30 times a second. Viewing a study one two-dimensional

[7]This research is by Robert Wilhelmson, Crystal Shaw, Matthew Arrott, Gautum Mehrotra, and Jeff Thingvold, NCSA

[8]This research is by Clint Potter, Rachael Brady, Pat Moran, NCSA/Beckman Institute



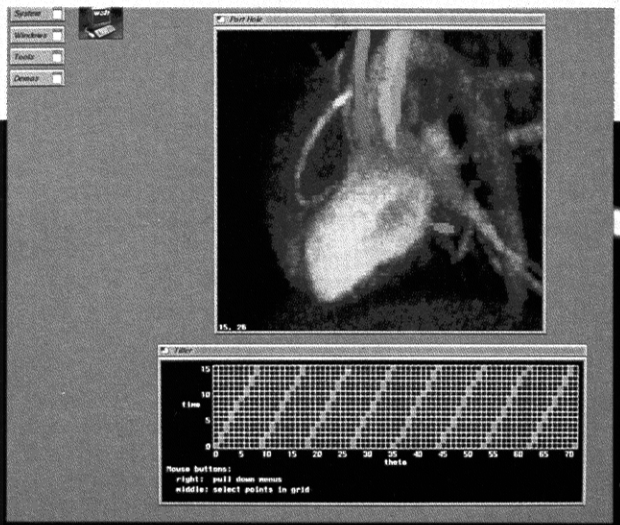


Figure 5.

Figure 6.

Figure 7.

plane at a time would take an enormous amount of time and still not directly reveal out-of-plane and temporal relationships.

The biomedical scientist requires computational tools for better navigating such an "ocean" of data. Two tools that are used extensively in the NCSA biomedical imaging activities are 'viewit' and 'tiller'.[9] 'Viewit' is a multidimensional "calculator" used for multidimensional image reconstruction and enhancement, and display preparation. It can be used to read instrument data, reconstruct, and perform volumetric projections saved in files as image frames. Each frame provides a view of the subject from a unique viewpoint at an instant in time. 'Tiller' collects frames generated by 'viewit," representing each frame as a cell on a two-dimensional grid. One axis of the grid represents a spatial trajectory and the other axis represents time. The user charts a course on this time-space map and then sets sail. A course specifies a frame sequence constructed on-the-fly and displayed interactively. This tool is particularly useful for exploring sets of precomputed volumetric images, allowing the user to move freely through the images by animating them.

At Showcase, interactive visualization of four-dimensional data will use an interface akin to that of 'Tiller'; however, the volumetric images will be generated on demand in real time, using the Connection Machine at NCSA. From a workstation at Showcase, the user will explore a large, four-dimensional data set stored at NCSA. A dog heart DSR data set from Eric Hoffman, University of Pennsylvania, will be used for the Showcase demo.

Figure 5. Scanning Tunneling Microscopy Laboratory at the Beckman Institute for Advanced Science and Technology.
Courtesy: Joe Lyding

Figure 6. Volume rendering sequence using "Tiller" to view dynamic spatial reconstructer data of a dog heart.
Credit: Pat Moran, NCSA

Figure 7. Three-dimensional rendering of Harvard CFA galaxy redshift data.
Credit: Margaret Geller, Harvard University, and NCSA visualization group

Scientific Multimedia Digital Library: The Scientific Digital Library[10] will be available for browsing and data analysis at Showcase. The library contains numerical simulation data, images, and other types of data as well as software. To initiate a session, the participant will use a Sun or SGI workstation, running the Digital Library user interface, to connect to a remote database located at NCSA. The user may then perform queries and receive responses from the database. The responses represent matches to specific queries about available data sets. After selecting a match, the user may elect to examine the data with a variety of scientific data analysis tools. The data is automatically retrieved from a remote system and presented to the researcher within the chosen tool.

One capability of the Digital Library was developed for radio astronomers. Data and processed images from radio telescopes are stored within the library and search mechanisms have been developed with search fields such as frequency and astronomical object names. This allows the radio astronomer to perform more specialized and comprehensive searches in the library based on the content of the data rather than simply by author or general subject.

The data may take the form of text, source code, data sets, images (static and animated), audio and even supercomputer simulations and visualizations. The digital library thus aims to handle the entire range of multimedia options. In addition, its distributed capabilities allow researchers to share their findings with one another, with the results displayed on multiple workstations—which could be located across the building or across the nation.

Navigating Simulated and Observed Cosmological Structures: The Cosmic Explorer[11] is motivated by Carl Sagan's imaginary spaceship in the PBS series "Cosmos," in which he explores the far corners of the universe. In this implementation, the user will explore the formation of the universe, the generation of astrophysical jets, and colliding galaxies by means of numerical simulations and VR technology. The numerical simulations produce very large data sets representing the cosmic structures and events. It is important for the scientist not only to be able to produce images from this data but to be able to animate events and view them from multiple perspectives.

Numerical simulations will be performed on supercomputers at NCSA and their resulting data sets will be stored at NCSA. Using the 45Mbit/sec NSFNET connection between Showcase and NCSA, data from these simulations will be visualized remotely using the VR 'CAVE'. The 'CAVE' will allow the viewer to "walk around" in the data, changing the view perspective as well as the proximity of the viewer to the objects in the data.

Two types of simulation data sets will be used. The first is produced by a galaxy cluster formation model and consists of galaxy position data representing the model's predicted large-scale structure of the universe. The second is produced by a cosmological event simulator that produces data representing structures caused by the interaction of gasses and objects in the universe.

[9]This research is by Clint Potter, Rachael Brady, Pat Moran, NCSA/Beckman Institute

[10]The digital library architecture and development work at NCSA is led by Jeff Terstriep.

[11]The Cosmic Explorer VR application software is based on software components already developed for VR and interactive graphic applications, including the Virtual Wind Tunnel developed by Steve Bryson of NASA Ames. Also integrated will be the software developed by Deyang Song and Mike Norman of NCSA for interactive visualization of numerical cosmology data bases, and the NCSA VR interface library developed by Mike McNeill.

Using the cosmic explorer and the 'CAVE', a user will be able to compare the simulated structure of the universe with the observed structure, using the Harvard CFA galaxy redshift database assembled by Margaret Geller and John Huchra. This will allow comparisons between the real and theoretical universes. The VR audience will be able to navigate the "Great Wall"—a supercluster of galaxies over 500 million light years in length, and zoom in on individual galaxies. Similarly, the simulated event structures such as gas jets and remains of colliding stars will be compared with similar structures observed by radio telescopes. The radio telescope data, as mentioned earlier, has been accumulated within the scientific multimedia digital library. This combined simulation/observation environment will also allow the participant to display time sequences of the simulation data, watching the structures evolve and converge with the observed data.

Acknowledgments

This article is in part an expanded version of the NCSA newsletter *Access* special issue on the metacomputer, Nov./Dec. 1991. Much information, text, and assistance was provided by Melanie Loots, Sara Latta, David Lawrence, Mike Krogh, Patti Carlson, Bob Wilhelmson, Clint Potter, Michael Norman, Jeff Terstriep, Klaus Schulten, Pat Moran, and Rachael Brady.

Further Reading

• Becker J. and Dagum L. Distributed 3-D particle simulation using cray Y-MP, CM-2. *NASA NAS, NASNEWS Numerical Aerodynamic Simulation Program Newsletter, 6,* 10. (Nov. 1992).

• Catlett, C.E. In search of gigabit applications. *IEEE Commun.* (Apr. 1992).

• Committee on Physical, Mathematical, and Engineering Sciences, Federal Coordinating Council for Science, Engineering, and Technology, Office of Science and Technology Policy. Grand challenges: High performance computing and communications. Supplement to the President's Fiscal Year 1992 Budget.

• Committee on Physical, Mathematical, and Engineering Sciences Federal Coordinating Council for Science, Engineering, and Technology, Office of Science and Technology Policy. Grand challenges 1993: High performance computing and communications. Supplement to the President's Fiscal Year 1993 Budget.

• Corcoran E. Calculating Reality. *Sci. Am. 264,* (Jan. 1991).

• Corporation for National Research Initiatives. 1991 Annual Testbed Reports. Reports prepared by project participants in each of five gigabit network testbeds.

• Hibbard W., Santek D., and Tripoli G. Interactive atmospheric data access via high speed networks. Computer Networks and ISDN Systems, 22, 1991, 103–109.

• Lederberg J. and Uncapher K. Towards a national collaboratory. Report of an individual workshop at the Rockefeller University, Mar. 1989.

• Lynch, D. Ed. *Internet System Handbook,* Manning Publications and Addison-Wesley, 1992.

• National Academy Press. Supercomputers: directions in technology and applications; [ISBN 0309-04088-4] Wash. D.C., 1989.

• *NCSA High-Performance Computing Newsletter.* NCSA's metacomputer: A special report. *access: 5,* 5 (Sept.–Dec. 1992)

• Smarr, L. Catlett, C.E., The Next Great Network *MIT Tech. Rev.,* July 1992.

• Stix G. Gigabit connection. *Sci. Am. 263* (Oct. 1990).

CR Categories and Subject Descriptors: C.2.4 [**Computer-Communication Networks**]: Distributed Systems—*Distributed applications, Distributed databases;* C.3 [**Special-Purpose and Application-Based Systems**]; I.3.2 [**Computer Graphics**]: Graphics Systems—*Distributed network graphics, remote systems;* I.6.3 [**Simulation and modeling**]: Applications; J.2 [**Computer Applications**]: Physical Sciences and Engineering; J.3 [**Computer Applications**]: Life and Medical Sciences

General Terms:

Additional Key Words and Phrases:

About the Authors:

LARRY SMARR is a professor of physics and astronomy at the University of Illinois at Urbana-Champaign (UIUC) and is the director of the National Center for Supercomputing Applications. His current research interests are in the investigation and visualization of astrophysical phenomena including black holes and the Higgs Field; email: smarr@ncsa.uiuc.edu

CHARLES E. CATLETT is associate director for computing and communications at the National Center for Supercomputing Applications, University of Illinois, Urbana-Champaign. He is developing applications and programming environments for the BLANCA gigabit/second testbed. email: catlett@ncsa.uiuc.edu

Authors' Present Address: National Center for Supercomputing Applications, 605 East Springfield Ave., Champaign, IL 61820