COMPUTING AND DATA PROCESSING PANEL

LARRY SMARR, University of Illinois at Urbana-Champaign, *Chair*
WILLIAM PRESS, Harvard-Smithsonian Center for Astrophysics, *Vice-Chair*

DAVID W. ARNETT, University of Arizona
ALASTAIR G.W. CAMERON, Harvard-Smithsonian Center for Astrophysics
RICHARD M. CRUTCHER, University of Illinois at Urbana-Champaign
DAVID J. HELFAND, Columbia University
PAUL HOROWITZ, Harvard University
SUSAN G. KLEINMANN, University of Massachusetts, Amherst
JEFFREY L. LINSKY, University of Colorado, Boulder
BARRY F. MADORE, California Institute of Technology
RICHARD A. MATZNER, University of Texas, Austin
CLAIRE ELLEN MAX, Lawrence Livermore National Laboratory
DIMITRI MIHALAS, University of Illinois at Urbana-Champaign
RICHARD A. PERLEY, National Radio Astronomy Observatory, Socorro
THOMAS A. PRINCE, California Institute of Technology
CHRISTOPHER T. RUSSELL, University of California, Los Angeles
ETHAN J. SCHREIER, Space Telescope Science Institute
JOHN R. STAUFFER, NASA Ames Research Center
GERALD JAY SUSSMAN, The MIT Artificial Intelligence Laboratory
JOSEPH H. TAYLOR, JR., Princeton University
JURI TOOMRE, University of Colorado, Boulder
JOHN A. TYSON, AT&T Bell Laboratories
SIMON D.M. WHITE, University of Arizona
JAMES R. WILSON, Lawrence Livermore National Laboratory
KARL-HEINZ WINKLER, Los Alamos National Laboratory
STANFORD E. WOOSLEY, University of California, Santa Cruz

Computing and Data Processing

I. OVERVIEW

Computational technologies have been central to advances in astronomy and astrophysics for at least the last four decades. Martin Schwarzchild's stellar evolution codes used roughly half the cycles of John von Neumann's MANIAC "supercomputer" during the early 1950's. The 1960's saw the first detailed supernova computations by Colgate and White. The Einstein X-ray observatory and the VLA radio array, coming on-line in the 1970's, created images by the use of computers as intermediaries between the sensor and the observer. Theorists used supercomputers to model a wide variety of complex astrophysical phenomena in the 1980's.

Outside of astronomy, on the national scene, the strategic importance of high performance computing to the future competitiveness of broad sectors of the U.S. economy is coming to be widely recognized. As a consequence, a major national focus (the High Performance Computing Program) is emerging. The proposed components of such a program include high performance computing systems, advanced software technology and algorithms, a National Research and Education Network, and support of basic research and human resources.

A national initiative in computing, whether the one now proposed or a different one, will usher in a new context in which scientific research of all kinds will be practiced. Astronomy in particular stands poised, by virtue of its intrinsic data- and computation-intensive nature, its manageable size as a discipline, its past experience and future opportunity, to be the cutting-edge application discipline in a number of major aspects of a national program. Astronomy's task is to build its own internal computer infrastructure, in such a way as to maximize its leverage *vis a vis* the national program – and simultaneously to bring to astronomers the computational technologies that will be enabling of innovative astronomical discovery.

Modern astronomy and technology are often inter-related. New developments in technology have spawned qualitative advances in astronomy, and the promise of scientific discovery has often pushed technologies beyond their existing state-of-the-art. The charge coupled devices (CCD), new technology telescopes, active optics, and computing technology are examples of areas currently rich in this synergism. Analog devices are being replaced in new instruments with digital devices based on digital signal processors with greater precision and stability.

On the observational side, the scale of astronomical data that will be gathered in the 1990s, and which must be manipulated, communicated, and archived, will be on the order of many terabytes per year. (A terabyte per week is perhaps a reasonable figure.) Interposed between observation and actual understanding stands, increasingly, multiple stages of highly intensive data processing. Operation counts in the teraflop range (10^{12} floating point operations) per reduced data set will be increasingly common. Teraflop

numerical simulations will be both helpful and practical in making the connections between astronomical observations,astrophysical theory and remote observing.

In the three complementary areas of digital data handling, intensive data processing, and theoretical modeling, astronomers are ready to take advantage of the expected technological advances of the 1990s: widespread of use of parallel computers, large increases in memory capacity, revolutionary improvements in data storage technologies, widespread use of graphics and visualization techniques, desktop high-performance workstations, high-speed networking, and powerful new algorithms.

The near future will see most researchers with access to powerful and flexible desktop computers linked over a national network, to each other as well as to high-value resources such as supercomputers and national observatories and data banks. Scientific visualization capabilities will be commonly available. The ability to bring together on the desktop the results of both complex simulations and detailed observations, and to be able to interact with each data set visually as well as quantitatively, could profoundly influence the progress of the astronomical sciences.

The Emerging National Information Infrastructure

In the 1960's, the Federal government provided the funds needed to set up first rate university computing centers. However, for fifteen years between 1970 and 1985, the Federal government removed itself from maintaining these facilities at the state-of-the-art. During that period few scientists had access to the newest computational technologies. Instead, shared departmental mini-supercomputers accessed by "dumb terminals" became the standard resource for most astronomers.

There was a radical reversal of this policy of "benign neglect" in 1985 when the National Science Foundation (NSF) formed the national supercomputer centers and began the national NSFNET network. These computational resources were financed from divisions of the NSF separate from disciplinary divisions. Access was not decided by money, but by peer review. Due to this democratization of access, in the last four years, over twenty thousand university scientists, engineers, social scientists, and humanists at over 250 universities and colleges have gained access to frontier computing technologies housed in the NSF supercomputer centers. There is a factor of 100 times the computing speed, memory, and storage capacity in the national centers as sits today on the desktop of the typical individual scientist. The National centers allow the benefits of substantial economies of scale with the cost of these facilities being borne across all fields of science and engineering. *We presume that the NSF, NASA, and DoE supercomputer centers, upgraded and enlarged, will continue to provide this resource to our community.*

During the same period, 1985–1990, individual workstations emerged which were as powerful as the previous departmental facilities. Most astronomers have managed to switch from "dumb terminals" to personal computers or workstations in the last five years. These desktop machines allow individualized control over one's computational research environment. The power and flexibility of these machines will continue to grow rapidly during the next decade. In addition, RISC (Reduced Instruction Set Computers) technologies have created a new version of the departmental computer which is near the speed and memory of a mini-supercomputer. The power of the departmental mini-supercomputers of the '90s will match or exceed those of the present generation of supercomputers. By the mid-1990's the computing power of the desktop computers, departmental minisupers and the central supercomputers will be at least 100 times what it is today.

The national network, which allows the researcher to "reach out" and grab that extra power when needed, has one thousand times the bandwidth compared to a user's access path just four years ago. The bandwidth of the national network will rise by yet another factor of 1000 during the coming decade. "Supernodes" arise naturally on the national network containing both specialized computational resources, and national digital archives of data - both from observations and from simulations. It is in computer networking that some of the greatest advances will come. As the gigabaud national network becomes a reality, there are three areas where revolutionary changes become possible. The first will be the use of facilities at the national centers from institutions all over the country. The second area is remote access to a distributed national digital library, which might contain scientific publications, previous observations, and results of theoretical simulations. Third is the remote control of "supertelescope systems" and the real time transport of the data to the astronomer.

We will also see a fundamental change in software. User interfaces are moving from command line, character oriented, single screens to menu driven, bit mapped, multiple windowed environments. It may come to pass that many important concepts in computer science (object oriented programming, distributed computing, and data structures) are becoming practical tools for computational scientists such as astronomers. A wide range of visualization technologies are changing the primary unit of information from number to image. Finally, large scale sharing of code is becoming accepted in at least the observational astronomy community, in the form of standard, portable, distributed data analysis systems.

The coupling of these elements in the next decade will transform astronomy and astrophysics into a *digital science*. The physical "glass and steel" of telescopes will be made orders of magnitude more powerful by addition of computing hardware and software. More researchers will "digitally observe" by accessing various national digital archives of all previously made observations, than will make new real observations on telescopes. Theoretical simulations of a complexity heretofore impossible will become commonplace. The ability to compare digital observation with digital theory will cross-fertilize both and lead to a much tighter mutual guidance than was possible in the past. Many of the aspects of this transformation will be shared with disciplines outside of astronomy, such as biology and environmental sciences.

Machines for the 1990's: Workstations to Supercomputers

Over the next decade, workstations will grow in performance to become comparable to present-day supercomputers. The present generation of RISC (Reduced Instruction Set Computers) -based machines are much more cost effective in terms of dollars per million floating point operations per second (MFLOP) than the current generation of supercomputers. These systems allow for affordable computing locally. Such systems also make possible a close coupling of analytic, numerical and visual computing.

Local systems offer several distinct advantages: First and foremost they offer the cheapest code cycle of any type of machine available. Second, the power and maintenance requirements (human and environmental) are much less. Third, and most important to the user, they offer i) instant response time, ii) little or no down time for maintenance, iii) local access for high resolution graphics, iv) sharing the resources with a much smaller group, as opposed to a central facility, where there are many times this number of users.

However, the advantages of supercomputers - large memory and disk capacity, vector or massively parallel processing, and extremely high input/output (I/O) rates, are crucial to a small fraction of computer users with large or demanding codes. A rough rule of thumb is the 80/20 rule: about 80% of users are performing small computations which can be supported effectively locally, *if* adequate funding resources are available. However, a small subset (perhaps 20%) of both the theoretical and observational community will attack problems whose computational requirements, in speed, memory, or storage, exceed those that can be reasonably provided locally. Such users will need access to national central facilities. The national centers will provide high cost technologies which will be available for experimentation by the community. These technologies will include vector processor and massively parallel supercomputers, very large memories, ultra high speed networks, large disk caches, the latest visualization technologies, all with teams of specialized experts.

The growth curves for desktop machines and for supercomputers are similar, so that with time, today's supercomputer capabilities will become affordable enough to be added to the local complement of machines. Of course, by then tomorrow's supercomputers will be more powerful too. Thus, the national supercomputer centers give researchers a chance to experiment with the future.

Essentially all active researchers need convenient access to good workstations, as well as "clear channel" coupling into the national network. In addition, small colleges and universities that have a history of training the students who become the future generation of scientists should be encouraged in their efforts to offer undergraduates exposure to modern computing. Support for high-performance workstations and mini-supercomputers would be a cost effective first step, by providing an accessible computational environment with modern software and graphics. It would also be a step toward geographical equality of resources and opportunity.

Lessons from the '80s

While certain subfields of astronomy (e.g., theoretical modeling) have always been demanding of

forefront computer performance, astronomy as a discipline has not, in the past, been consistently out in front of other fields in the computing arena. However, during the 1980's the breadth of the computational base within astronomy has been expanding, a number of remarkable developments have begun, often from small beginnings, which are now poised to bring qualitative changes to the discipline.

A key example is the use of image processing algorithms in both optical and radio astronomy. These methods have long been used in radio astronomy, where synthesis arrays do not themselves form an image, but depend upon a digital computer as the image forming element of the telescope. However, much more than a simple Fourier transform of the measured visibilities is now standard practice. Powerful deconvolution algorithms have been developed which can greatly enhance the power of both radio and optical imaging telescopes. As discussed in the "Array Telescope Computing Plan", a conceptual proposal submitted by the National Radio Astronomy Observatory (NRAO) to the National Science Foundation (NSF) in September 1987 and resubmitted recently, the original design goal of the Very Large Array (VLA) of a dynamic range of 100:1 has now been increased routinely to 2,000:1, and a dynamic range of 100,000:1 is achieved for point sources. Thus, the VLA can be seen as an evolving telescope, with today's version being an instrument orders of magnitude more powerful and flexible than the one which was designed – all without hardware design modification. But the cost of this extra power is in the computing.

Unfortunately, the NRAO has insufficient computer power available to allow the full potential of the VLA to be realized. The computing problems with the VLA data have two origins - the first is the sheer volume and the second is the processing speed. To quote the above proposal: "The 10% of the expected proposals that generate 70% of the computing workload ... will be processed ... in supercomputers at national and regional centers. The rest will simply be deferred (i.e. users will not schedule the telescopes realizing that computational resources are not there or they do not reduce the data they have)..." .

In many cases, observations which comprise the most exciting and innovative of the possible radio synthesis projects cannot be carried out for lack of computing resources. This minority of projects is of great astronomical interest. We can summarize the types of scientific projects which lie within this category: a) All low-frequency imaging. In order to allow proper imaging at low declinations, VLA is designed as a 2-D instrument. This then requires a 3- D imaging for large fields of view. Due to special problems inherent in the computing requirements, all imaging at frequencies lower than 1 GHz, and much of the data taken at 21cm and 6cm must be processed with three- dimensional transforms. This is discussed in detail in section IV. b) All snapshot programs. One of the unique features of the VLA is its ability to make 2-dimensional images of bright, compact objects in only a few minutes of observing. This ability results in a speed enhancement of up to a factor of 200. But the cost is in computing. c) Studies of the interstellar medium of individual galaxies. These require extremely large images (up to 4096 × 4096 pixels) with 128 velocity channels. d) All Galactic absorption studies. These also require large images (comparable to the above) with high velocity resolution. e) OH and H2O maser emission. These are again, large images with high velocity resolution. The large VLBA computing needs are dominated by the same types of spectral line projects as listed above. All of the projects listed above except item b) require three-dimensional imaging from very large data bases. In some cases, four dimensional hyper-cubes may be required.

One can view the NRAO experience either as a great success – hugely multiplying *solely by processing* the peak capabilities of an expensive national instrument – or as a cautionary failure, a failure of vision (or national resource allocation) to provide the necessary computer power to a premier national astronomical facility. From either point of view, the conclusion is the same: astronomers are now sensitized to the importance of powerful computers and powerful algorithms, and they are determined that the negative aspects of the VLA experience will not be repeated.

The VLA experience is one of the best-documented examples of *computer starvation.* However, it is not the only example that we might offer. The Infrared Astronomy Satellite (IRAS) threw away a vast amount of information by binning its data too coarsely, due to computer hardware limitations. (This is now being redone by the Air Force!). While the Infrared Processing and Analysis Center (IPAC - the IRAS archive, administered by Jet Propulsion Laboratory on the CalTech campus) has been widely praised for its accessibility and servicability, it nonetheless has noteworthy limitations imposed on it by the computing power available to it. Another example involves the EINSTEIN observatory database. Although a unique resource for more than a decade, lack of adequate funding (until very recently) constrained the database and software to the mid-1970's technology on which it was developed. The forthcoming Gamma Ray

Observatory (GRO) data system also has significant computational shortcomings. The coming VLBA and the expanding millimeter-wave arrays such as the Berkeley-Illinois-Maryland Array will once again raise the threat of computational resources limiting and directing the science which can be done. On the positive side, the Hubble Space Telescope data analysis system (STSDAS) was designed to be exportable to a variety of computers. Furthermore, the budget for observers and archival researchers has been protected, and this funding can be used by successful, peer-reviewed proposers in part to acquire computer resources as appropriate. NASA has also recently encouraged work-station procurements in other projects, although lack of fundseasily compromises the program.

The technology trends of the next decade can dramatically improve this situation. It is important that our *national observatory system*, coupling multiple remote user analysts with data-gathering telescope facilities (both ground-based and space-based) have resources allocated to track the trends in computing technology. Emphasis must be placed both on a distributed computing environment, developing software to run efficiently on existing architechtures, including the difficult to program but extremely powerful massively parallel systems and making time available on existing supercomputers.

In light of the foregoing, our findings for astronomical computing are straightforward. Resources must be available for individual workstations, and for departmental or observatory mini-supercomputers. Networks must link the desktops of all investigators, all observatories, and all data archives. The development and maintenance of community software assets such as national data archives, data analysis programs, and theoretical simulation codes should be fostered. The allocation of computing resources is best carried out by peer review, but some oversight by the field is necessary to assure balance. The context for these findings is the assumption that current support from non-astronomical funding sources for the national network and the supercomputer centers continues throughout the decade.

Arrangement of This Report

Section II reviews some major challenges and technology trends encountered in facing the transformation to a digital astronomy - on both theoretical and observational grounds. Section III sets out in detail the need for a national data archive, and discusses some of its dimensions. Section IV consists of four "case studies" of high-performance data processing (both observational and theoretical), each one attempting quantitative estimates of what the requirements in the coming decade will be. Section V discusses how the transition from today's Megabit/sec national network to a 1990's Gigabit/sec fiber optic net will alter both observations and theory.

II. THE TRANSFORMATION TO A DIGITAL ASTRONOMY

In this section, we briefly review how astronomy and astrophysics will gain considerably from the technology trends and the implications of the national information infrastructure. With a discussion on real-time data processing, remote observing, theoretical simulations and community code development efforts, we show how our discipline is well poised to provide a leadership role in bringing the transformation and infrastructure into existence.

Supertelescopes

An emerging viewpoint is that all observational or laboratory instruments are *"smart sensors"* – a coupling of detectors to computers. The scientific power of a modern telescope is greatly leveraged by the amount and sophistication of the computing hardware and software applied to it. The lesson of the VLA is that a telescope is no longer a fixed capability instrument. Rather, it becomes a "supertelescope" which becomes more powerful with time, by virtue of its coupling to new generations of more capable digital computers. The balance of the "silicon to steel" tradeoff in designing a multiple decade national astronomical facility must be taken much more seriously in the 1990's than it was in the 1980's. Our report focuses attention on several examples of this, including large field CCD optical-IR arrays and radio telescopes. The increase in the sensitivity and resolution which computers will add to telescopes can be comparable in importance to the construction of large new telescopes now under construction or in the planning stages.

The NRAO, has proposed a distributed approach to meeting the computing challenges mentioned in the last section: 1) By writing and supporting flexible, all-purpose, exportable code (i.e., AIPS), much or even most of the projects scheduled on the VLA can be properly reduced at the observatory on mini-supercomputers or at the researcher's home institutions on desktop or departmental computers. This approach satisfies perhaps 85% of the individual observing runs, but falls far short of providing the capacity for the few demanding projects. 2) To provide the capacity for very large projects, the NRAO advocates a supercomputer access plan. The required software, perhaps specially coded to match various high performance architechtures, would be available on a few very large capacity machines. Fast data links must be available to allow real-time interaction of the user with the results. This is required since so much of radio astronomical data reduction is iterative in nature, and an experienced eye is required to judge the progress.

An idealized scenario for remote creation of VLA supermaps might work like this: The user physically or electronically sends his or her data sets to a designated contact at the national center. This person arranges to load the data onto disk. The user accesses the data on the supercomputer through their home workstation. The required commands can be issued from home, and the incremental results can be quickly transferred back to the workstation through fast data links for viewing by the user. After a number of iterations, which might take from hours to days to complete, the final results can be permanently archived, and the data deleted. Obviously, an efficient management structure will be needed to make this work. And, user-friendly, familiar code must be available to support the remote user.

A different high performance computing challenge faces optical/IR observers. Large charge-coupled device (CCD) focal plane imagers in the next generation of instrumentation for very large ground-based telescopes will require pre- processing in near real-time. Cameras with mosaic detectors larger than 5000x5000 pixels are now possible. The data rate from these detectors will overwhelm not only the traditional mini- or micro-computer or workstation, but also current array processors attached to mini- computers. About 1Gb of raw images (mostly calibration data) would be acquired in each 24hr period per instrument. Routine recording of such volumes of raw data for later reduction and analysis would create a data bottle-neck which would prevent the science programs from being carried out effectively. Real-time automated preprocessing and initial analysis of these data will be required. Special processors are now being built which can handle the high data rates from such large CCD detectors.

The scientist does not know precisely what is in the data, nor that it can be analyzed in one pass. What is necessary is real-time preprocessing of the raw data through all processing steps which are proven and which do not sacrifice other interesting scientific data. Reduction of data volume by a factor of at least 10 would result, for both imaging and spectroscopic CCD data.

Past examples of real-time array processing can be found in the fields of remote sensing, mail sorting, process inspection, radar signal processing, underwater topography, medical imaging and machine vision. Massively parallel real-time processing is constrained by the problem of transferring parallel data from a serial data stream at sufficiently high data rates. As in biological systems, analog image preprocessing at the detector becomes an advantage. Analog charge-coupled computing for focal plane image processing has been implemented in experiments. Neural networks, particularly analog VLSI preprocessors, have applications in real-time image processing. Digital computers have continued to keep pace with developing imager technology, so that most existing and planned astronomy instrumentation data systems are digital after the detector output and A/D converter. Eventually, optical computers are expected to increase array processing speed a thousand-fold.

Because optical telescopes form images directly, the optical community has been slower to experiment with deconvolution processing of their images than the radio community. However, the maximum entropy method (MEM) can be fruitfully used on CCD frames obtained under relatively poor seeing. Recent comparison of the results with frames obtained under better seeing, have shown that MEM deconvolution can improve the effective seeing substantially without creating any spurious features or structure. In addition to the effects of seeing, deterioration of images due to poor guiding and diffraction effects of secondary supports can be corrected. Routine MEM processing of all images taken with a particular telescope would require something like minisupercomputer power. The rapid rise in the processing power per dollar of available computers will within the next few years make it feasible to have a dedicated high-performance computer as an integral part of every optical telescope used for the acquisition of astronomical images.

Observing from your Desktop

Currently, many scientist's valuable times are inefficiently used in travel to remote sites to operate telescopes during observing runs. In addition, the optimal scheduling of a telescope is impossible, for the program of the astronomer on site always has the priority. If, instead, an astronomer at his or her home institution could monitor the data as it was obtained remotely with the same data rate as he or she would have at the site, and if telescopes could be dynamically scheduled as weather or other conditions change, the utilization of both the telescopes and the productivity of astronomers would increase.

The high performance national network will provide for *"teleobserving"* – the remote control of telescope systems and the real time transport of the data to the astronomer. In most cases the data will be obtained automatically, in several observing sessions, over a period of time. In other cases, observers will be notified if they wish to observe interactively. This will allow us to address the problems of optimizing the use of scarce national observing resources.

In this environment, the distinction between space and ground- based observing will begin to disappear. NASA's Great Observatories will automatically observe lists of targets under software control. New-technology ground-based telescopes, with their multiple fixed instruments, will use an optimum observing strategy, making rapid observing mode changes possible. Thus, for the optical-IR observer, a mode similar to that now used on the VLA may become common. Queue and "program" observing, taking optimum advantage of changing atmospheric conditions, will obtain the best possible data for all projects. These developments have the potential to change the way most observers work.

Electronic communication is also needed for real-time operations. For instance, planetary astronomy runs "campaigns", which may be multiple wavelength studies of the same object coordinated in time. They could be centered on a stellar occultation or mutual eclipses of a planet's moons. Such campaigns benefit from tight communication among the observers. Furthermore some planetary phenomena have time scales shorter than the terrestrial rotation period so that worldwide networks of telescopes are needed to properly characterize them.

Finally, the network can allow some synthesis radio telescopes to operate in near real-time. The present operational mode for synthesis radio telescopes is to acquire data and store the data on magnetic tapes for off-line processing at a later time. There is little opportunity for the astronomer to immediately see the results of his observations while the telescope is still available for follow-up observations - needed either because of poor data quality or to follow up an exciting, unexpected result. Particularly with a radio synthesis array telescope, the interval between data acquisition and working with the processed data ranges from weeks to infinity.

As a testbed and prototype of a tightly coupled telescope system, high speed network, and supercomputer, the Berkeley-Illinois-Maryland Array (BIMA) plans to implement a near real-time radio telescope. The BIMA will be a six-antenna millimeter-wave array with eight separate spectral windows of 256 channels each being available. In a typical eight hour tracking, the visibility function will be sampled sufficiently to make useful data cubes (8 cubes of 8 different spectral lines with right ascension, declination, and velocity axes) immediately. The telescope system, physically located at Hat Creek, California is completely under computer control and is accessible via computer networks, so an astronomer can monitor in real time the data acquisition process, edit the data, and set up command files for data processing. The data will be sent from Hat Creek to Berkeley to the supercomputer at the University of Illinois for immediate calibration, mapping, and deconvolution using an "expert system" controlled by the command files set up by the astronomer. An astronomer in Berkeley will then be able to display and begin the analysis of the data cubes on a local workstation by using a gigabaud national testbed network. If problems are found with the data, the project can be scheduled for immediate reobservation while the telescope system is still in the same configuration; if something unexpected is found, new observations can begin immediately.

This tightly coupled system of telescope, network, supercomputer, and workstation will very significantly raise the utilization of telescope systems and productivity of astronomers, and if successful, will be a prototype for all modern supertelescopes.

The NASA astronomy community is studying similar approaches, especially in the context of future space and lunar-based observing. Here, the need for automation in mission planning, expert systems for data analysis and experiment monitoring, space-borne data processing, advanced data compression and communications technology take on added significance. NASA is sponsoring studies in these areas, and is

being encouraged to involve the user community in prototyping these technologies. Again, without direct user involvement in both the prototyping and all phases of the eventual implementation, the new capabilities run the risk of being inappropriate for the desired purposes.

Astrophysics in a Numerical Laboratory

Because of the nature of our (primarily) observational science, astronomers seldom can actively probe the objects of interest. Often these objects are complex in both form and temporal behavior, which hinders theoretical description even in cases in which we have the correct ideas regarding the underlying physics. From the beginning of the development of the digital computer, astrophysicists have been using this tool to simulate complex observed systems and to experiment numerically with new theoretical concepts.

Astrophysics depends on theory and modeling to a greater degree than other physical sciences, because astronomers can only observe remotely; active experimental intervention is not possible. Moreover, the observed phenomena are typically the result of usually complicated interactions among highly nonlinear processes occuring simultaneously. Therefore, it is necessary to construct rather elaborate models to achieve a satisfactory interpretation of the observations. As a consequence, the photons and fast particles which escape from astrophysical objects must be theoretically analyzed to the hilt, to extract meaningful physical information about the nature of their sources.

There is a long tradition of using analytic simplified models to capture the essence of a complicated astrophysical phenomena. Desktop computers are becoming increasingly important to support this work. Modern symbolic mathematics software allows the theorist to use more complex analytic formulations of the problem. Ordinary differential equations, which required a supercomputer to solve in the 1960's, are routinely evaluated and graphed by workstations and personal computers today. During the next decade the power of desktop machines will become so great, that many of the problems for which astronomers and astrophysicists are using supercomputers today, will also become soluble locally. Thus, we believe that desktop computers have become absolutely essential for theoretical astrophysicists.

As researchers build more and more complexity into their models, they outstrip the ability to compute locally in a reasonable turnaround time. This complexity arises for two fundamentally different reasons. First, the spatial dimensionality grows from one to two to three. Furthermore, as more and more realistic models are attempted, systems which are first studied as static, become time dependent. Typically, as the geometric complexity grows, so does the number of physical variables which must be solved for (e.g. from a radial velocity vector in spherical symmetry to all three components of the velocity vector in general). Second, one adds additional physics to the problem which increases both the number of equations and their coupling. For instance, one may add magnetic fields, nuclear or chemical reactions, radiation transport, or viscosity to an inviscid fluid flow code. In some cases, the introduction of new physics raises the effective dimensionality of the problem. For example, to describe the radiation flow in the most general case, one would add two angle variables and one frequency variable to the calculations. If there are scattering terms in the sources (Compton, Thomson, Rayleigh etc.) the system to be solved is a seven-dimensional integro-partial-differential equation. In addition, the calculation of realistic properties (opacities, equation of state) strain the resources of current machines almost to the breaking point. In short, both geometric complexity and additional physics can rapidly drive up total computational time and memory to values far exceeding today's fastest machines.

Real astrophysical systems are 3-dimensional objects evolving in time with extremely complex physics. Some aspects of these systems are currently being simulated on today's workstations, supermini computers, and supercomputers, subject to the restrictions on physical and geometrical realism which are imposed by the user's computer hardware and software. The goal of software designers is to make it possible to run codes transparently on any computer on the network, while retaining the interactivity and familiarity of local facilities.

The 1990's will be the decade where a number of long-standing astrophysical problems will be solved and computers will play an important role in these solutions. Areas which seem particularly ripe for rapid theoretical progress, and comparison with observations, can loosely be categorized as follows: large scale structure of the universe and cosmology, active galaxies and jets, star formation and the interstellar medium, dynamics of stars and stellar atmospheres, supernovae, accretion onto compact objects, generation of gravitational radiation, and the microphysics and magnetohydrodynamics of astrophysical plasmas. We

will see significant advances particularly through modeling and numerical simulations approaching realistic complexity, which can be directly compared to observations.

As more powerful computers and community simulation codes become available, incorporating realistic physics and, where necessary, full three dimensional time-dependent geometry will greatly increase the ability of astrophysicists to directly compare their simulations with observational data. To illustrate the sort of progress we expect, consider a typical problem where interactions of radiation with matter are crucial. During this decade we will see the addition of nonequilibrium physics to hydrodynamics codes. A few codes have already taken a step in this direction with the introduction of two or three temperature systems comprising, say, electrons, nuclei, and radiation. But none yet allow for nonequilibrium effects in the excitation and ionization distributions. When this is done, radiation field and the state of the material become inextricably interwoven, making it impossible, even in principle, to calculate the thermodynamic properties of the material in terms of purely local variables. Rather, the system becomes fundamentally nonlocal, and we are forced to solve very large systems of globally interlocked equations, characterized by a wide range of characteristic spatial and time scales. These problems require the resources of massively parallel machines, and we should devote considerable efforts to algorithm development for such machines, forming an effective alliance with computer science experts working on such machines.

We estimate that nearly 10% of practicing astronomers are presently engaged in theoretical simulation of astrophysical phenomena. Some of this computational astrophysics is being done using local workstations and mini-supercomputers. Of the total allocations of time for all areas of academic science and engineering on the NSF Supercomputer Centers facilities, roughly 10% of the resources are being used by researchers in the field of Astronomy and Astrophysics. This is equivalent to about three processors of a current generation supercomputer which would cost around $20M to purchase. Those who are using supercomputers are trying to solve problems that push the system to the limits of software and hardware capabilities in existence today, and which could not be addressed using local computing resources. Some of these projects are also the ones attacking key problems in the discipline, and making seminal contributions that lead to major paradigm shifts in astrophysics. It has to be recognized that before the establishment of the national supercomputer centers, it was extremely difficult for astronomers to gain access to supercomputers. Consequently relatively few students were trained in the use of these machines, and the number of actual users remained very low. Now that the national centers exist, it becomes practical, for the first time, to train students in computational astrophysics. We just now have the first generation of these students receiving degrees and becoming professional astronomers. The percentage of professional astronomers who will be carrying out large computational simulations will grow rapidly over the next decade (assuming that the national supercomputer centers remain adequately supported).

In conclusion, the computing hardware needs of the theoretical astrophysics community can, with certain important exceptions, best be filled by a distributed system consisting of local mid-range computing facilities, including super-mini computers and graphics workstations, and upgraded national supercomputer facilities and high-speed network links.

Community Software

In order to effectively utilize the enormous advances in computer hardware expected in the next decade, we must have an accompanying development of scientific software. This is actually more costly and should be of at least equal concern with the computers themselves. Code development activities often require tens of man years, followed by a sizable budget and group for their maintenance.

The observational astronomy community has shown an admirable degree of coherence by developing systems like AIPS (Astronomical Image Processing System) and IRAF (Image Reduction and Analysis Facility), which have been adopted widely. These packages have saved an immense amount of time and duplicated effort. However it has often been difficult to identify adequate funding for ongoing efforts in the crucial areas of code maintenance and modernization. It is critical to augment efforts in the latter two areas.

From the perspective of theoretical simulations, code development efforts pertaining to multidimensional hydrodynamics codes, magnetohydrodynamic and particle-in-cell plasma codes, advanced stellar structure and supernova codes, N-body codes etc. should be encouraged and funded. At least part of this effort might fruitfully be located at a supercomputer center or a national laboratory, because this type of institution has

the required broad infrastructure, and houses many related activities which are synergistic with the types of software development needed for astrophysics. Modest funding here might enjoy large leverage.

The astronomical community has an excellent record in the definition, maintainance, and distribution of uniform software platforms. Indeed, in the area of image processing, it appears that astronomy has already taken a technical leadership position relative to other scientific disciplines. Standard software development is a vital activity for the health of the commmunity: without the distribution of such software, and the use of standards, the handling of digital data is expensive and inefficient – or else doesn't get done at all.

One big success has been the FITS (Flexible Image Transport System) data format, now used internationally for the exchange and archiving of astronomical data. FITS was developed in 1979 by NOAO, NRAO, and NFRA (Netherlands Foundation of Radio Astronomy), and is an openly published standard. Since 1982, FITS has been the IAU standard for data interchange in astronomy. The FITS standard is maintained by regional committees in Europe and North America, which act under the authority of the FITS Working Group under Commission 5 (Astronomical Data) of the IAU. Recently NASA has established the FITS Support Office inside NSSDC at Goddard Space Flight Center.

The development by the NRAO of the Astronomical Imaging Processing Sysytem (AIPS) software illustrates both good and bad features of the software of the last decade. AIPS provides a very functional image processing system, with standards maintained by a national center. The general image processing capabilities of AIPS are sufficiently powerful that AIPS has been used extensively for optical and infrared image processing. An attempt was made during AIPS development, to isolate machine dependent features.

Software development for new large projects can learn from the experiences of the earlier models – a small core of people dedicated to the development of maximally transportable and evolvable software with the widest possible distribution in, and contributions from, the community.

The model introduced by NRAO for radio astronomy has also been adopted by other disciplines in the observational astronomy community, particularly by the large optical and x-ray groups. The National Optical Astronomy Observatories (NOAO) started developing (in about 1980) the Image Reduction and Analysis Facility (IRAF), a portable data analysis system designed to support their user community, and the European Southern Observatory (ESO) also started developing the Munich Interactive Data Analysis System (MIDAS). The development of AIPS, IRAF, and MIDAS provides the community with a limited number of very functional astronomical data analysis systems, which are portable to many computing platforms (ranging from PCs to minisupercomputers), and distributed widely, with standards maintained by national centers. The advantage of such a coordinated approach is demonstrated by the general willingness of the astronomy community to adopt these systems. STScI adopted IRAF as the environment for the HST data analysis system (STSDAS), and SAO adopted a similar approach for the ROSAT system (PROS). There are thus now several functioning groups, associated with national-level facilities, creating standardized software environments for data manipulation, analysis, and display. Grass-roots coordination has evolved among these groups, as well as via AAS and IAU working groups. The main impediment to further progress in this area would appear to be the lack of adequate funding, especially for maintenance of generic capabilities, not required for a specific project. Nonetheless, in the area of data analysis and image processing, it appears that astronomy has already taken a technical leadership position relative to other scientific disciplines.

The future will see more use of open systems, and standards, and very high speed networks. It will be essential that modern modular software standards be followed and that software be written to be portable to a variety of computers and usable over national high speed networks. Both the the national observatories and national supercomputer centers must lead in the astronomical software development effort. An example of this software effort is the development of a method for storing and transferring multi- object files across different machines on the network. One wishes to keep multi-dimensional floating point data arrays, palettes, images, and annotations together under one file name (for instance, in observational data sets such at FITS files, or theoretical simulation data sets). Further, one doesn't want to have to bundle and un-bundle these objects by hand. These computer science constructs should be discipline independent. An example is the Hierarchical Data Format (HDF), developed by NCSA. HDF allows different vendorsU computers on the network to automatically access combined files. The user's application code can read and write HDF files. NCSA is working with the NSF and NASA national observatories to create translators from the discipline specific file format FITS to the discipline independent file format HDF.

Careful attention to the "lifecycle" of software is also necessary. Major space missions such as the Great

Observatories are being designed for an operational lifetime of 15 years and for use by a large number of observers and archival researchers over a period of roughly 25 years. Given the typical 10 year development cycle, the ground data systems for these missions must function in a cost efficient manner over a 35 year period, despite rapid change in computing hardware, operating systems, and data analysis packages. Of even greater importance to the astronomical community is the very large cost for the ground data system of the first Great Observatory, the HST, and for subsequent missions unless fundamental changes are made. Over the lifetime of a long-lived Great Observatory, the ground data system may become as expensive at the construction costs for the observatory, and the fixed amount of money available for mission operations and data analysis may lead to a situation where rising costs of the ground data system will decrease the funding available for astronomical research.

One important way to reduce the costs of ground data systems is to design them from the start to accommodate rapid change in computing hardware and operating systems. Thus these systems should be built with evolvability and portability of software as requirements. In particular, layering to provide independence from specific operating systems and hardware is highly desirable. This design philosophy may be more expensive initially, but it will be very cost effective over the long term of these missions. Ground data systems should be portable to the major data analysis packages like IRAF and AIPS, which execute on a variety of vendor's platforms.

The HST Science Operations Ground System (SOGS) is often used as an example of the old methodology of developing ground systems, where a set of both operations and user requirements were implemented via a major formal procurement, and resulted in a large, monolithic, vendor-specific, hardware/software system. In fact, the portions of the ground system which were developed relatively late in the project, either with significantly greater user involvement, or by the users themselves, tend to be more in keeping with current ideas on evolvability and portability, as well as being more responsive to user needs. In the area of planning and scheduling, a portable expert system (SPIKE) has been added; in operations, workstations which run IRAF have been added to support off-line analyses and displays; and the pipeline calibration processing utilizes the identical algorithms available to any researcher through IRAF/STSDAS.

The end users of data from telescopes on the ground and in space are the people best able to determine sensible requirements for the software systems that will process the data. For this reason they should play a major role in formulating the requirements for such software systems, and they should closely monitor the development and testing of these systems.

This section has concentrated on observational astronomy community software. This is because such software is probably about one decade ahead of community codes for theoretical simulation. Although there has always been informal sharing among computational astrophysicists, there are not many truly national community astrophysics codes. Other theoretical fields such as chemistry, electron device simulation, plasma physics, and engineering have a rich history of the use of such community codes.

Efforts are underway at the national centers to develop, distribute, and support national users with new application software for astrophysical fluid dynamics research, incorporating the most accurate algorithms available for modeling astrophysical fluids. Versions will be developed for dynamics in 2- and 3-spatial dimensions, incorporating the important physical effects of self-gravity, magnetic fields, radiation, and thermodynamic properties of the gas. This software will incorporate the most accurate algorithms available for modeling astrophysical fluids in the Newtonian regime. The goal is an evolving software package implemented in a modern distributed UNIX operating system environment and optimized for high performance computers. Other environments needed include workstation tools with user-friendly interfaces for pre- and post-processing.

In addition, software for performing *"numerical observations"* of the simulations will also be developed. Numerical observations refer to the process whereby the fundamental physical variables of the simulated model are translated into observables (intensity, line widths, line shifts, polarization angle, etc.), including an assumed instrumental response, so that direct comparisons with observations can be made.

III. THE NEED FOR A NATIONAL ARCHIVE: SHOULD WE CONTINUE TO THROW AWAY DATA?

Historical Considerations

The first astronomical archives were just the written recordings of what astronomers saw in the sky. Since no detectors were involved, little was lost in the process, to the extent that the astronomer described his visual impressions in sufficient detail. These archives were quite useful to later astronomers. In general, these archives were not considered proprietary and were shared with other interested parties when that was requested. The archival medium proved quite durable; a variety of recent papers on sunspot numbers, supernovae, comets, etc. were based on observations many centuries in the past.

As photographic plates became suitable for astronomical purposes, and instrumentation developed to utilize this new type of detector, plate vaults became the new astronomical archive. Plates proved remarkably suited to astronomical demands, and remained the detector of choice for optical astronomy at least into the late 1960's. These plates have been, and will continue to be, extremely useful as a storehouse of information for the study of stellar motions, objects which vary in luminosity or morphology, objects whose spectra vary in time, etc. The degree of availability, or even records of the existence, of plates of a given object or region of sky varies greatly from observatory to observatory and observer to observer. However, a considerable fraction of the photographic database is available to the persistent astronomer.

Beginning in the 1960's, electronic detectors began to replace photographic plates for an increasing number of uses. By the present date, the main role remaining for photographic plates is wide field imaging, where the large format size of the plate compensates for the much lower quantum efficiency relative to two-dimensional electronic array detectors. The development of these electronic detectors, unfortunately, acted as a deterrent to maintaining a data archive. Problems with the data from the new detectors include: (1) the detectors have evolved rapidly in time, and proliferated in variety, making it difficult for an outsider from the future to understand what is available sufficiently well; (2) the data reduction required for the electronic detector data is often arcane, and specific to the detector and program being conducted, with an imperfect record of the required procedures; (3) the formats for storing data on tape or disk were generally not standardized, were in some cases machine dependent, and were often not well documented; and (4) the media on which the data were stored have, in many cases, become obsolete, been lost or overwritten, or deteriorated over time so that the data have been effectively lost. Thus, for some types of archival research, the recent electronic era may represent a regression - future astronomers will not be able to reconstruct the evolution of some phenomena because the relevant observations were not archived in a reliable fashion.

Problems and Opportunities for Archiving Ground-Based Observations

Some of the problems described above for data derived from electronic detectors are inherent to the field. Despite these problems, there are a number of reasons for optimism that now is a good time to seriously consider large scale archiving of ground- based data. It is important to note that space-based astronomy data is now routinely archived by NASA policy. Although subject to many of the aforementioned problems, it is nonetheless possible to retrieve most data obtained in the last 1.5 decades. In the case of the Great Observatories, standard data formats and standard data analysis systems are the norm. The development of the FITS format for data transfer and storage has allowed astronomers anywhere in the world, using a variety of computers and operating systems, to read and write tapes that are readable anywhere else. Given the will and the resources to smoothly transition to other storage media, FITS provides the means to create a long-lived astronomy archive so that data from the present will be accessible to future astronomers. Another important recent advance is the development of multipurpose, machine (and vendor) independent data analysis packages - such as AIPS, IRAF, and MIDAS. These packages provide standardized, well-documented methods to analyse a large percentage of the types of data produced with ground-based telescopes. In optical astronomy, CCD's have become the dominant detector for many purposes, and have thus become - at least temporarily - a de facto standard, further simplifying the task of archival research. In the planetary sciences community, with many of the problems similar to that of the astrophysics community, archiving is done with the Planetary Data System. Finally, but quite importantly, the recent development

of computer networks and the imminent implementation of the NASA Astrophysics data archive system provides the means and the model for allowing widespread access to archives of ground-based observations.

Recent technological advances make this a good time to be considering development of an archive system. In the last few years, use of computer networks by astronomers has increased dramatically. The throughput of the system has improved considerably, and can be expected to improve much more in the near future. From a much different direction, the development of very large 2-D array detectors (2048^2 CCD's in the optical - soon to be mosaiced, and 256^2 arrays in the infrared - also soon to be mosaiced) makes data archiving more important for two reasons. First, the larger arrays mean that there is a greater chance for serendipitous discoveries - larger areas of the sky, or larger portions of the spectrum, make it more likely that data obtained for one purpose may contain information useful for another. Second, the greatly increased volumes of data make it much more likely that the data will not be fully utilized - valuable information relevant to a number of projects might be brought back from an observing run, but the PI only has the time and manpower to address the one question for which the data were obtained.

The fact that it is physically possible to develop an archive system for data from ground-based telescopes does not justify actually doing so. There must be an expectation that the scientific return for developing such a system is greater than the return for spending an equal amount of money to obtain new data. We believe that in many cases, the argument can be made in favor of archiving. CCD images obtained today provide an irreplaceable source of material for later proper motion studies. Similarly, the search for supernovae progenitors and other objects where one wishes to search for variability require access to archival data. During the next decade, a large number of US and international space missions will discover myriads of new and interesting sources - if we have had the foresight to be archiving ground-based data, much valuable information on these new sources may be obtained without using any new telescope time. Data archives that are well-documented and easily accessible over networks will provide fast "food for thought" following brain-storming sessions (instead of "... that's a great idea. Let's write a telescope time proposal, and wait six months ...", it can be "that's a great idea. Let's find out if anyone has obtained an image of that region in the last five years ...").

Impediments to Establishing a Data Archive for Ground-Based Astronomy

Data obtained from ground-based telescopes have traditionally been the property of the observer in perpetuity. Establishing proprietary periods and development of data archives would break this tradition. Many arguments could be advanced to keep the status quo: (1) the present system protects astronomers pursuing long-term projects; (2) graduate students should be allowed to finish their theses without having their data become public property; (3) astronomers with heavy teaching loads may not be able to reduce their data rapidly; (4) only the person who took the data fully understands it, and making it available to the masses would create more confusion than benefit; etc. We believe these are not reasons against archiving, but are instead just implementation questions - the rules determining propriety periods must be flexible enough to allow for differing circumstances, and the archival system must include calibration information and other relevant procedural data.

A more serious argument against archiving data from ground-based telescopes is simply that there is no money. This is particularly true for NSF funded national observatories such as NOAO and NRAO. The implementation of the archive system would require additional work. Archiving of data in Astronomy must be put into perspective with the other (time and resource) competitive aspects of doing our science. Observing with ground- and space-based telescopes, reducing the data, and publishing the results derived from them, all require time, expertise and money. Like everything else, archiving should be driven by the quality of the science that it delivers. Given limited financial resources, open competition and peer review - an explicit and case-by-case review for funds to archive the data is the most sensible approach. Our legacy is not just a stream of binary data, to be cared for and then preserved for future generations. Our legacy includes a distillation of those data into a perspective on our field that drives future thought, brings about a synthesis of new ideas and then precipitates innovation and additional observations. While we certainly cannot second-guess what future generations will find of interest in our presently gathered data, at the same time we cannot premise our science on that self-same lack of information; we cannot save everything out of ignorance. Justification as to the inherent value must be made competitively, proof that archiving will conform to the standards (of formatting, documentation, history etc., for raw data, calibrated data,

and fully reduced data) must also be given and later verified. None of these additional tasks should be so onerous as to outweigh the potential benefits.

The contrast with NASA's outlook and expenditure for archiving and archival research is marked. NASA astrophysics currently spends of order $5 million per year strictly for archival research, primarily for data from IRAS, EINSTEIN and IUE. This amount is likely to grow considerably during the next decade as the archives for HST, ROSAT, and AXAF become available. STScI, working with other astronomy groups, has developed a prototype optical disk archiving facility which currently incorporates its digitized sky survey, and is being used to archive the early HST data. NASA is developing a major new large scale archive and distribution system for HST data which is estimated to cost over $20 million. A considerable amount of money will also be spent developing the Astrophysics Data System (and Master Directory) which will link the various NASA astronomy data archives together. Meanwhile, essentially no money is available to archive ground-based observations. The US observatories are lagging behind a number of major foreign observatories - archive programs have been started at the Anglo- Australian Telescope, the European Southern Observatory, and the La Palma Observatory.

Compatibility of Ground-Based and Space-Based Archives

Along with an archive there is a need for a catalog of the archive. This catalog should be as complete and easy to use as possible. Data and archiving systems for astrophysics space and ground-based missions should be compatible with, and ideally be a part of, the Master Directory Service of NASA's Astrophysics Data System. The data reduction and analysis software should be available at users' home institutions, not simply by remote access over networks to the mission data center. In particular, the available I/O bandwidth (over networks) would severely limit user access to the system. In contrast, the user community has (or soon will have) sufficient computing power locally to process the data. Finally, the scientific community should have free access to the archive.

It is desirable to develop cost-effective, useful archives of digital data from ground-based astronomical observations, available over a high-speed national network. Data are a legacy which we have a duty to bequeath to our successors. We note and commend NASA's important steps in this direction in the arena of space-based observation.

We recommend a similiar initiative for ground-based data with the following specific goals: (1) All major ground-based observatories, both public and private, should incorporate the capability for archiving of digital data. (2) The archived data should be accessible over the national network along with backup mechanisms such as rapid mail delivery of massive datasets on appropriate media. (3) This archive should consist of major homogeneous data sets with their requisite calibration information. Observatory Directors should establish and oversee appropriate criteria for the implementation of archives. We recommend that the funding agencies include the archiving program when determining the allocation of resources to an observatory. (4) The NSF should support high quality proposals for funding the capital costs of archiving data obtained at private observatories, and these data should become public after a proprietary period that extends no longer than 18 months after the last data in an observing program are obtained. (5) Archives must be designed to outlive any specific hardware, software, or media. (6) Archives should include all the raw data, calibration data, and information necessary to remove instrumental signatures from the data. (7) All observations obtained with large ground-based telescopes should be catalogued, whether or not the data are placed in an archive. (8) Original observers should get their data in the archive format so that both the original and subsequent analyzers of the data will start with the same data set. (9) A policy should be established to archive data described in papers published in the refereed literature. This policy should be enforced and implemented through journals, time-allocation committees, and the proposal reviewing process. The AAS Council should take actions to facilitate the archiving of processed digital data concurrently with publication in the Astrophysical Journal and the Astronomical Journal. (10) The standard for data interchange should be FITS or FITS extensions approved by the IAU and the NASA FITS Standards Office. Standards should be developed for data compression, archiving, catalogs, and user interfaces. (11) Peer review is essential for determining the allocation of resources for archiving in the context of all other competing requests for resources within astronomy.

We recommend a study of mechanisms for community input and user review of archive plans and operations; integrating ground-based archives with NASA's Astrophysical Data System; how further interagency

collaboration on archiving between NASA and NSF should be implemented; and the impact of archiving plans in Europe and elsewhere.

The issue of international collaboration and coordination in archiving, and networks should be addressed. Reciprocal agreements must be reached with other international agencies with regard to access to and support of data archives. Astronomy on an international scale will develop much faster than in the past due to network access. Also a large percentage of the total useful astronomical database will be located outside the U.S.A. NASA has already begun to address this question with missions such as GINGA, EXOSAT and ROSAT, but no standard solutions exist; rather case-by-case solutions have been adopted. A goal should be to achieve a universal approach to archiving and data analysis. NSF should approach its foreign counterparts concerning ground based archiving and data access.

Giant model calculations are sometimes as valuable as observations. The numerical experiments together with their codes are directly analogous to major observational data together with the details of their acquisition techniques and calibration data. The implementation plans for archiving should include data from numerical computations with the source codes becoming public after an initial proprietary period.

Digital archiving will become increasingly important in sciences other than astronomy. Early serious study, and commitment of funds, can put astronomy community in a strong position as a testbed for other disciplines. This would allow NSF astronomy to seek joint funding with other Divisions, or from NSF management and Congress directly.

IV. HIGH-PERFORMANCE DATA PROCESSING: OBSERVATIONAL IMAGES AND THEORETICAL SIMULATIONS

This section presents four independent "case studies" – individual visions by panel members of the computational frontiers of astronomy in the 1990s. The first case study attempts to put a quantitative scope on the needs for realistic theoretical simulation. The second describes the specifics of one particular case, namely plasma astrophysics. The third case study documents the information explosion that is about to occur in optical data collection, with the advent of large CCD detectors and their attendant processing requirements. The fourth study is a corresponding analysis for the case of radio images.

Case Study A: Realistic Dynamical Simulations of Complex Systems

The "Fourth Dimension Supercomputer" is a system sufficiently powerful to calculate the evolution of complex, nonlinear systems in a fully three-dimensional space and in time. Presently large memory supercomputers are barely capable of providing this capacity, but only with crude resolution. Present trends in computer technology suggest that the next generations of machines will be serious "Fourth Dimension Supercomputers". To see what this implies for our science, let us examine what would be the parameters of a "Fourth Dimension Supercomputer", and what are the needs associated with its use.

Parameters needed for three dimensional simulations

Let us take a measure of our understanding to be how well our concepts match the images were observe. What is the quality of a "good" image? A variety of arguments and painful experiences with more meagre values suggest that a moderately good image requires about 500 x 500 pixels. The fact that in the personal computer market a popular screen resolution is about this same value is probably no accident (and 1000 x 1000 would be better). Our images are projections of 3D objects onto a 2D surface, so that a numerical description of the objects itself would imply $500^3 = 1.25 \times 10^8$ points. How many of these visualization points can be represented by a single computational point? This depends upon the power of our computational algorithm; significant advances are being made in algorithmic development. If several pixels are sufficient to describe a computational point, we have about 200 computational points per spatial dimension. Experience suggests this resolution is low but beginning to be interesting for a variety of problems involving simple physics, such as inviscid gas dynamics. We will assume $200 \times f$ to be required number of computational points per spatial dimension in subsequent discussion. This may be thought of as a minimum acceptable resolution for a variety of problems of current interest; $f = 2$ or $f = 3$ may often be appropriate. At each

point a 3D calculation would require 4000 or more floating point operations (e. g., multiplications, additions, etc.), giving a value of $3.2 \times 10^{10} f^3$ "flops".

What about the resolution in time? Consideration of accuracy and stability of the numerical method often results in requiring a number of time steps which is of the order of the number of space steps in one dimension – in this case $200f$. For ten natural times, a short but not uncommon value, this implies $10 \times (200f)^4 = 1.6 \times 10^{10} f^4$ points, and $6.4 \times 10^{13} f^4$ floating point operations. A fast workstation may now provide 1 Megaflop performance, so that (if it would fit in memory) this task would take $6.4 \times 10^7 f^4$ seconds, or about $740f^4$ days. The most efficient code on a four-cpu supercomputer might run at 400 Megaflops, so that the task would take $44f^4$ hours. For twice the linear resolution, $f = 2$, this increases to 30 days.

Memory

How much memory is required? A floating point number of full precision requires 8 bytes. The physical state of a single point in the system may require 20 or more numbers. For two time slices (an old and a new state), this implies $3.2 \times 10^8 f^3$ numbers or $2.56f^3$ Gigabytes. The largest memory presently available to astrophysicists at NSF supercomputer centers (1 Gigabyte) is about $2.5f^3$ times smaller. Consequently on present supercomputersthe task would have to be paged in and out of memory, with attendant problems for speed and storage resources.

Storage

Using this value of $1.28f^3$ Gigabytes per state (time slice), and assuming that 100 of the $2000f$ time steps are saved for analysis, the storage needed per project is $128f^3$ Gigabytes. For 40 such projects, the storage requirement grows to 5 Terabytes. This is a major limitation: the extensive storage at the NSF supercomputer centers for researchers in all areas of science and engineering is estimated to be of the order of a few Terabytes. Data compaction and new storage technologies are needed to alleviate this bottleneck.

Communications

How do we get these data to the scientist for analysis? A reasonable estimate may be obtained by requiring that the time for data transfer of the results must be less than the time for computation — otherwise the data flow becomes the bottleneck. For one session of 100 time steps calculated, a 400 Megaflop supercomputer requires $8 \times 10^3 f^3$ seconds. It generates a new state needing $128f^3$ Gigabytes of storage. This implies a data tranfer rate of greater than 1 Megabit/second, which is well above the actual performance of Ethernet, for example. Data compaction and upgrade of the NSF backbone to T3 are needed. It is particularly important to provide this national network resource to the wider community of scientists and students who do not reside at a supercomputer center site.

Speed

While low resolution projects are feasible on present supercomputers, higher resolution places serious demands upon computing speed. For 40 research groups per supercomputer, there are 200 hours per group per year (efficient parallel use of all cpus is assumed). At $44f^4$ hours per project, this allows 4 projects per group per year, which is not really adequate for even moderate surveys of parameter dependence and testing. As progress is made on the constraints above, there will be demand for greater cpu speed, especially as observational comparisons make higher resolution necessary. Note that even this estimate is optimistic, as it assumes a level of vectorization and parallelization of code which is only occasionally obtained in practice.

Local Storage

Most scientists and students are not at the supercomputer sites. Upgrading the networking system will allow data to flow to them, but there must be local facilities to deal with it when it arrives. For example, a project of $2000f$ time steps, generating $20f$ saved states of $1.28f^3$ Gigabytes each, would require $2.56f^4$ Gigabytes of local storage. Factors of 2 and 4 could be saved by using 32-bit and 16-bit wordlengths.

It seems inescapable that analysis of 3D computations requires interactive 3D viewing, and that in turn requires extensive local storage capacity.

Viewing

Analysis of dynamical 3D systems requires looking at the time behavior of the system. This implies a requirement for local graphics capacity. It should be noted here that local RISC based machines are capable of high quality visualization. For example, suppose we want 60 seconds of images at 10 screens/second. For 3 Megabytes/screen (3 byte color on a 1000 × 1000 screen), this is 1.8 Gigabytes needed on a high speed disk to feed the graphics engine. This graphics pipeline must be fed at 30 Megabytes/second.

Algorithmic development in the area of 3D imaging is rapid, but needs support for standardization. Otherwise there will be a lot of redundant development of very similar software.

Case Study B: Plasma Astrophysics

Over the next decade, sophisticated numerical models and simulations will play a particularly critical role in the field of plasma astrophysics. The reason for this lies within the intellectual structure of the field itself.

It is widely supposed that plasma-physical mechanisms are responsible for many of the non-thermal processes observed in astrophysics, such as high-energy particle acceleration and the coherent emission of radiation. Similarly, non-classical transport mechanisms, such as anomalously large viscosity in accretion disks, or anomalously high resistivity in astrophysical dynamos, seem to be required by current astrophysical models. Plasma-based processes are at the heart of the micro-physics of these transport phenomena. An important goal for theoretical astrophysics is to develop quantitative calculations of the expected nature of these plasma processes, and of their observational consequences in relevant astrophysical situations.

But plasma processes both determine, and are determined by, their parent system's global configuration. Experience with laboratory and space plasmas has shown that a plasma's behavior is sensitive to the specific physical conditions and geometry in which it finds itself. At the same time, some knowledge of the plasma's behavior is often essential to constructing a credible large-scale model of the astrophysical system in question. Thus in order for astrophysical plasma physics to produce quantitative results that can be meaningfully related to astronomical data, an iteration must be performed between the microphysics (simulations of microscopic plasma processes) and the large-scale configuration which emits the photons observed by astronomers (simulations via hydrodynamic models, usually either radiation- hydrodynamical or magneto-hydrodynamical).

The field of plasma physics has been a pioneer in the development of successful computational models, including descriptions at the kinetic, magnetohydrodynamic (MHD), hybrid, and fluid levels. Indeed this development has been a necessity, due to the nonlinearity and geometrical complexity inherent in collective plasma behavior. Fusion research, of both the magnetic and laser-driven variety, has made extensive use of computational simulations in the interpretation of data from laboratory experiments, helped by the facilities of the National Magnetic Fusion Energy Computer Center and at national laboratories. Likewise, NASA's support of large-scale computing within the solar- terrestrial theory program has made computational simulation a regular tool for interpreting in situ space-physics data from NASA's solar- system probes. Driven by the fusion and space-physics communities, the computational simulation of microscopic plasma processes has shown considerable success over the past decade.

Unlike laboratory or space plasmas, one cannot probe the conditions in astrophysical plasmas directly. Thus astrophysical plasma physics research must take the additional step of integrating microphysics models with appropriate large-scale system models, so as to arrive at a quantitative prediction of the observed photon output. A start in the direction of large-scale astrophysical models has already been made. In the field of solar physics, MHD studies of turbulent convection and fluid-magnetic-field interactions will allow detailed comparison with the next generation of high-resolution solar instruments. Similarly, the first generation of MHD models of astrophysical jets has reached a sophisticated level, allowing comparison with high-resolution radio data. In addition to the further development of these two areas, over the next decade one can anticipate the development of MHD models for the large-scale structure of accretion disks, supernova remnants, pulsar magnetospheres, solar active regions, and planetary magnetospheric structure.

An important feature of the next generation of macroscopic system models will be the incorporation of results from detailed plasma simulations at the micro-physics level. For example, non-linear transport coefficients developed using small-scale plasma simulations will be used within larger macroscopic system models to predict the photon output. Similarly, source terms for non-thermal or relativistic particles can be developed using plasma simulations, and then applied when the appropriate conditions emerge in a large-scale macroscopic model.

Computational Requirements: Microscopic Plasma Simulations

Most state-of-the-art microscopic plasma simulation codes are currently being run on multiprocessor vector supercomputers, particluarly if more than one spatial dimension is involved.

What does the future hold? Mini-supercomputers will be used increasingly for the less demanding simulations. At present it is not clear whether massively parallel architectures will be well-suited for particle simulations of plasmas except in some special cases, although they may be useful for some types of Vlasov or hybrid algorithms. However, overall there are strong pressures towards moving to next-generation multiprocessor vector supercomputers. The reason for these pressures lies in the need to push beyond the very small volumes that can presently be studied using microscopic plasma simulation methods, and in the need to perform three-dimensional simulations in order to study geometrically complex phenomena such as magnetic field line reconnection. Thus plasma astrophysics has genuine need for supercomputer resources of the class that the NSF and DOE Centers can potentially provide.

Hand-in-hand with the need for supercomputers is the need for advanced graphics and visualization capabilities to interpret the results. Many microscopic plasma simulations follow the evolution of the distribution function of electrons and/or ions in phase space, together with gradients in real space. Thus, present kinetic models are frequently 4- or 5-dimensional (two space dimensions and 2 or 3 velocity dimensions), and future models will add a third space dimension as well. Advanced visualization techniques will be a prerequesite for extracting useful information from simulation models having this high level of complexity.

Computational Requirements: Macroscopic System Models

Simulations which incorporate the results of plasma micro-physics studies into a model of the large-scale astrophysical system have a slightly different computational flavor, although many of the computational requirements are similar to those described in the previous subsection.

In magnetohydrodynamic (MHD) models, only two or possibly three dimensions are involved. Thus it is possible that the memory and speed requirements of these models can be met using the present and next generations of mini-supercomputers, coupled with the type of advanced graphics and visualization tools described above. Massively parallel architectures are also a possibility for future MHD models, although much research remains to be done to optimize performance in this area.

However in the end what is useful for astrophysics is a prediction of the radiation output. Thus some sort of treatment of radiation emission and transport will be a critical element of many macroscopic system models. Once radiation transport is added to a fluid or MHD model, the number of effective dimensions increases, taxing the memory and speed capabilities of (at least today's) mini-supercomputers. Likewise, radiation transport introduces coupling in angle or in frequency which is difficult to treat on massively parallel architectures. Thus the use of state-of-the-art supercomputers will be critical for this type of macroscopic modelling effort.

Case Study C: CCD Optical Images and Image Processing

Large charge-coupled device focal plane imagers in the next- generation instrumentation for very large ground-based telescopes will create massive amounts of data. Real-time automated preprocessing and initial analysis of these data will be required. In the past, image processing in astronomy has generally not been on-line or real-time. The correlators on the VLA radio telescope are a good example of automated data pre-processing, but the image data on that telescope are not automatically processed.

Progress in automated photometry of crowded fields was made with the software packages DAOPHOT and ROMAPHOT. In the radio, the AIPS package contains several semi-automated routines for cleaning. Recently, near real-time preprocessing of digital images has been made possible by standardization of image header information and advances in processor/storage hardware. Data in the image header may be used as process history and keys. That is, stored images (flat field, bias, dark, object exposures) required in the processing may be retrieved automatically by reference to key entries in their headers, including date, time, and filter/spectrograph settings. The IRAF/CCDRED package permits automated pre- processing of large volumes of raw 2-dimensional images, and is gaining popularity in the optical community.

Automated Image Analysis Software

Image databases have been large (several Gb) in optical astronomy for years. The automated detection, classification, and photometry package FOCAS (Faint Object Classification and Analysis System), developed over the last ten years, has enabled statistical studies involving large image databases. FOCAS is a collection of image analysis and automated pattern recognition programs designed for automated reduction and morphological classification of astronomical images. Recent FOCAS releases (now available as part of the IRAF package) include powerful image pre-filtering operators along with interactive color-graphics display programs which allow the user to quickly identify objects with selected properties, such as color or two-dimensional shape.

A 100 million pixel imager

Current silicon CCDs cover about 1% of the quality imaging area in the focal plane of large telescopes. It is now necessary and possible to construct a mosaic of CCDs which cover a larger focal plane area. Let us consider a 5x5 mosaic containing 25 CCD arrays, each 2048x2048 pixels. The peak raw data rate for the camera would be 100 Gb/night. It is nearly possible, by current 1990 technology, to process, reduce, analyze, and archive this imaging database in nearly real-time.

A CCD mosaic imaging survey of a 100 degree patch of the sky would produce vast amounts of data which would have to be managed and processed. The resulting detected object rasters in several wavelength bands, and the FOCAS matched catalog, would comprise 600 Gb. The data rate from these detectors will therefore overwhelm the traditional mini- or micro-computer or workstation. Recording the raw data, without any on- line preprocessing and display, is cost-ineffective: about 100Gb of data would be acquired in each 24hr period.

It is imperative that the design of large CCD imagers make adequate provision for easy and rapid data analysis. The magnitude of the data processing tasks required for this mosaic imager would require a special-purpose system. The characteristics of this system are dictated by the requirement for real-time image correction and automated analysis, but the same hardware would be capable of performing extensive image post analysis.

The design of the data system must emphasize computational power, fast data transfer paths, flexibility, and expandability. Almost as critical as processing the raw data, this would provide the astronomer with a powerful workstation for exploration of the reduced images and catalogs. Due to the need for instant access, image displays, and interactive image analysis, remote supercomputers are not a solution to the computational requirements in optical-IR image processing and analysis. Instead, we look to fast processors based on Digital Signal Processors with GFLOP average speed and multiple wide busses, together with large RAM memory and fast multi-port disks, which are now becoming available.

Consider a multi- wavelength digital sky survey using a 10^8 pixel imager. The final CCD images for each band are passed to the FOCAS automated detector and classifier creating a catalog of properties (isophotal, aperture, and total magnitudes, centroid positions, and several central moments) for each detected object. For this mosaic imaging survey, with a limiting magnitude < 22 magnitude, the sky is sparse and the resulting catalog would be very small (a few Gb) compared with the full processed image databank (8000 Gb). The FOCAS point spread function is automatically determined from the stars in the image. Detection proceeds by convolving the image with a slightly broadened point spread function, and demanding that a real object must have more than say 10 simply connected pixels above 3 times the rms convolved sky noise.

Terabit digital archives

New technologies are emerging which will allow archiving of the resulting large image and catalog databases. Tera-bit optical recorders are now available. If the total imaging survey reduced data (8000 Gb, mostly blank sky) were stored on 2.4 Gb optical disks, it would require 3300 disks, close to $1M. Since access time is not critical for archiving, another technology is very appealing: optical film. Spot density of one per micron and areas up to 1 inch x 2000 feet may be obtained cheaply. Over several hundred Gb of encoded digital data may be archived on such a medium. Hardware for recording and reproducing in this format exists. These recorders can sustain 3 Mb/s data rates. Both the multi-band image rasters and the FOCAS catalogs (less than 600 Gb total) could be archived inexpensively.

It would even be possible to store the entire 8000 Gb in a collection of optical tape reels no larger than a feature length motion picture. It is clear that it is not practical to save the data and analyze it later, an approach which is already causing some problems with the current generation of small CCD cameras: the data simply piles up. To avoid this analysis bottleneck the images must be corrected as they are acquired. The CCD mosaic imager would produce a continuous data stream of 4 Mb / sec. One night's observing would typically produce 100 Gb of data. The final mosaic image is the result of extensive mathematical corrections applied to this data stream. During this correction operation each 1 Gb of data may move from disk to memory and back several times. Mis-alignment of the CCD rows in the mosaic would also be corrected in this processing.

In summary, we will soon have mosaic imager/computer systems capable of pushing the largest existing telescope to its performance limit. High efficiency CCD imagers covering most of the useable focal plane, together with specialized on-line computers using automatic image classification software will radically alter our ability to observe the universe.

Case Study D: A "Typical" Large VLA Data Processing Request

It might be difficult to grasp the enormity of the computing problem for VLA data without an example given in some detail. Most of the computer-limited problems are three-dimensional in nature, usually from spectral-line data. One special problem - imaging low-frequency data, is continuum in nature, but nevertheless requires the spectral line observing mode. Below we describe this computing problem.

The NRAO has recently completed installation of 327 MHz receivers on the VLA. Unique science addressable with this new capability includes steep spectrum objects and objects of large diameter and low surface brightness. Due to the two-dimensional geometry of the VLA, the samples of the visibility function are made throughout a three-dimensional volume, and the conditions under which a two-dimensional Fourier transform can be used to recover the source brightness fail, requiring much more expensive solutions. The simplest of these is a three-dimensional Fourier transform, producing a three-dimensional image 'volume' whose axes are in direction cosines, and within which the desired image is found on a sphere of unit radius. Processing of the image follows the same procedures normally done in two-dimensional processing – for example, deconvolution proceeds in an entirely analogous fashion using the three dimensional image with a three-dimensional beam.

In low-frequency imaging it is necessary to process all the sources within the primary beam. To reach maximum sensitivity, all confusing background sources must be located and removed through deconvolution.

A typical large project at 327 MHz will use all four VLA configurations with perhaps 12 hours observing in each. Because of chromatic aberration, the spectral line correlator must be employed to ensure the bandwidth of each data channel remains small. The result of this is a 16-fold increase in data rate and volume over the continuum case. In this mode, radio frequency interference, can be identified and purged without seriously corrupting adjacent channels. The integration time for each sample must be kept very short to prevent time-averaging smearing. The result is a very large database: Typically 3.5 GBytes, containing over 500 million complex numbers.

Calibration of these basic data is straightforward, and can be accomplished with modest computer resources, providing only that the short-term disk space to contain the data is present. These data must then be written to tape, or other storage medium, perhaps optical disks or high density tapes, with 1 - 2 GBytes.

The imaging needs are exceptionally large. The simple 3-D transform requires making a "dirty" map

and beam, each $4096 \times 4096 \times 64$ pixels, which with 4 bytes per pixel requires 8.5 GBytes memory if they are made in the most efficient, straightforward way. Fortunately, another approach is more efficient in memory, although at the cost of I/O and CPU. The image can be built up through a series of a large number of subimages of limited depth. The memory requirement is relatively more modest, about 135 MBytes per facet.

Deconvolution of the image is the next concern. In the simple, single cube approach, the procedure is straightforward, although highly consumptive of memory. No recall of the data is required. In the polyhedron imaging approach, deconvolution can be accomplished with much reduced memory, again at the cost of much increased I/O. A rough estimate is that perhaps 10,000 Fourier transform subtractions of clean components from the data are required, along with 10 re-imagings of the entire field. That is, more than 2500 FFTs, each 1024×1024 will be required. But this does not finish the processing, since the ionospheric corruptions must be removed through self-calibration. Self- calibration and deconvolution are interlinked, the former using the results of the latter to generate a better estimate of the sky brightness. Typically, three loops of self-calibration and deconvolution are required before satisfactory convergence has been achieved. Thus, all the operations described in the previous paragraph must be multiplied by three.

A rough estimate of the time required can be made: Using the Cray Research Inc.'s CRAY-2 supercomputer performance with a more modest case as a benchmark, and multiplying up by the ratio of database sizes and number of fields to process, results in a rough estimate of 250 DAYS for full processing. The time required is dominated by the gridding: Each visibility point must be distributed over about 100 adjacent cells, and averaged with all other visibility values within this cells, resulting in a computation-limited problem. We are confident that useable short-cuts will be found, as detailed studies of solving this computing problem have barely begun. For example, we can probably use a much less expensive gridding algorithm, which could cut processing by a factor of five.

It seems clear that the optimal approach will eventually involve massive parallelism. We could imagine 16 parallel machines, each comparable to, or faster than the CRAY-2 supercomputer. The data are then distributed to each of these machines, each of which is responsible for one sub-field. A central processor will be required to handle the component model subtractions – the model comes from all 16 parallel processors. Factoring in these expected savings, and imagining future, more powerful machines, predicts this particular problem to be soluble with a few hours of computing time.

The essential points of the above example are summarized here: First, there is a need for very fast machines with very high I/O rates and extremely large memory to generate the data cubes. In many cases, parallel processing is clearly advantageous. Second, assuming the NRAO will not be able to obtain such machines, the national supercomputer centers must provide access and support for astronomers requiring this computing. Third, home machines, such as workstations, must be supported to allow proper interaction between astronomer and completed image. And fourth, research and development of computing algorithms must be actively supported, both at the national center(s) and at the NRAO. The latter site is particularly important, for only at the observatory are the problems fully understood, and the vested interest present on a daily and continuing basis – factors which are absolutely essential to ensure progress in imaging science.

The revised VLA computing plan will handle most of the large VLA requests. But perhaps in 10 years, the algorithms and computers will allow the local computing environment to handle even these large requests.

V. NATIONAL HIGH PERFORMANCE NETWORKING: OBSERVATIONAL IMAGES AND THEORETICAL SIMULATIONS

In this section, we describe a major national experiment in networking, just getting underway, with a goal to determine how remote users will be able to interact at high speeds with remote supercomputers, observatories, and digital archives. The Corporation for National Research Initiatives, is organizing a set of five national gigabaud testbeds, which will become an integral part of the High Performance Computing Program. One of the testbeds will be transcontinental in scale and will have as application drivers an astronomical observatory and a distributed dynamical 3D simulation. This testbed, called NRI Blanca Testbed, will provide a first look at how the high performance computing infrastructure of the 1990's will enhance theoretical and observational astronomy.

The NRI Blanca National Gigabaud Testbed

The NRI Blanca National Gigabaud Testbed will create a prototype distributed scientific laboratory involving researchers at a number of universities on a fiber optic network. The plans for the transcontinental testbed network is to start with 45 Mbit/sec rates currently, with a goal of approaching a Gigabit/Sec over the next five years. Supercomputing facilities, large scientific databases, and high- performance visualization workstations will be connected via this Gbit/sec network, with data collection and observatory sites, and with collaborating researchers at each site. Research projects which involve information exchange in the form of data sets or interactive images or both - with volumes that definitely require a network running these speeds, will be supported.

Additional development efforts to be included in this project involve laser disc technology archive systems, image generation algorithm development, and development of a fully distributed, general purpose scientific simulation control and visualization system. The distributed visualization and simulation control system will be of general use, with libraries and client/server processes which can be used by computational scientists regardless of the specific discipline involved.

Simulations and image processing on supercomputers often require access to data bases at remote sites which are too large to be moved and/or which are being collected at high rates. Further, programs running on the supercomputers must be controlled by researchers from remote locations, requiring visualization output at that remote site which is a) of high resolution such as is necessary to determine the accuracy and quality of the run, and b) displayed in real time to allow control of the supercomputer application process.

BIMA–A High Performance Computing Observatory on the Gigabaud Testbed

Future supertelescopes will have as an essential component a very high speed data link between the sensor and a computer. Real-time radio astronomy would revolutionize the field by permitting an observer using a synthesis array to see an image of the radio sky as the observations were being made. Interactive observing could be a reality if the image processing can be done and the images transmitted fast enough. The goal of the BIMA experiment on this national gigabaud testbed is to demonstrate such capabilities and to explore how such capabilities might improve, expand, and extend the power of a telescope system.

The Berkeley-Illinois-Maryland Array (BIMA) is located at the Hat Creek Observatory in northern California and is operated by the University of California at Berkeley, the University of Illinois, and the University of Maryland. The Array is similar in concept to the VLA, but operates at millimeter rather than centimeter wavelengths. By early 1991 BIMA will consist of 6 antennas; there are plans for expansion to as many as 12 antennas. The BIMA system has been chosen for this testbed because the proposed gigabaud network will extend from Berkeley to Urbana, linking the sensor with the supercomputer, and because BIMA will generate data and have computational needs which are a significant fraction of those of the VLA. Although with 6 antennas BIMA has only about 5% of the number of simultaneous interferometers as the VLA, the BIMA spectrometer produces 4 times as many spectral channels and allows observation of up to 8 spectral lines simultaneously; the density of spectral lines in frequency space at millimeter wavelengths means that much of the time this multiplexing capability will be employed usefully. Further, BIMA will be used in spectral-line mode essentially all the time, while the VLA is often used in continuum mode. The BIMA data rate and computational requirements will be about 1/3 those of the VLA. A gigabaud connection between Socorro, New Mexico and one of the supercomputer centers would allow, in principle, similar remote operation of the VLA and the VLBA.

A typical BIMA data set will be in the 100 MB to 1 GB range; such data sets can be transferred from Berkeley to the supercomputer at Illinois at 45 Mbaud (real and sustained) in the period of a 5 minute coffee break. The initial processing of the observed visibility data on the supercomputer will be automatic, under the control of an "expert system" with tunable parameters which may be set in advance by the astronomer. While the observations are in progress, calibration, map making, and an initial deconvolution, self-calibration, and mosaicing (if appropriate) of a partial data set may be carried out and the data cube returned to the astronomer at Berkeley for analysis on a workstation. The astronomer will be able to judge the quality of the data, to see if the signal is strong enough to proceed with the observations, to judge whether the area of sky being mosaiced is correct, and to begin to experiment with processing parameters. Instrumental or atmospheric problems can be detected quickly, and corrections made or re-observations

carried out while the telescope is still in the same configuration. Exciting or unexpected results can be pursued immediately. When the project observations are complete, the full data set can be processed interactively on the supercomputer from 2000 miles away. The processing of radio maps is often highly iterative and interactive. The astronomer in Berkeley will be able to examine each step in the deconvolution (CLEAN or MEM) and self-calibration process as it runs on the Cray and fine-tune the algorithm parameters to yield the best possible maps.

Today, such interactive observing is possible only for astronomers in Urbana and only to a limited extent, because of the slow speed of the shared NSFNET. Using the recently developed MIRIAD (Multi-channel Image Reconstruction with Interactive Analysis and Display) software, the sizes of the images which can be processed can reach 4096 x 4096 pixels, images referred to as supermaps. Data are loaded into a supercomputer, which processes it and sends images of the processed data to a frame buffer connected via HPPI (High Performance Parallel Interface) at up to 800 Mbit/sec. This allows the local researcher to observe the calculations in real time; changing parameters and regenerating images interactively. Today, the local researcher can send 2-3 1024 x 1024 x 24 bit images per second (this being the resolution of current display hardware), which allows direct interaction with the image processing of quadrants of supermaps.

During the next 5 years the network capacity will allow 4096 x 4096 x 24 bit images to be transferred in under 0.5 seconds per image, which will enable the same level of interactivity remotely on full "Supermaps" as is available today at Illinois on a single quadrant; MIRIAD can transfer the desired 2-3 images per second and still maintain total interactivity with the image reconstruction.

Combined with applications for multiple simultaneous viewing by separate workstations (multiple collaborators located at multiple remote sites) this will allow for a level of interactive collaboration which is not feasible today. Combined with systems such as the digital archive for astronomical images, this network rate will allow for paging through multiple images. The remote researcher will have the ability to process much of the existing raw data which has not yet been viewed or evaluated over the network.

Remote Control of Fourth Dimension Supercomputers

Tools will be developed in the gigabaud network testbed project to build applications which support real time collaboration among multiple, remote scientists on scientific and computational aspects of a simulation running in real time on a supercomputer. The specific application chosen as a platform with which to demonstrate these tools is the study of storms using a four dimensional numerical model (3 space dimensions and time). The tools developed in this national testbed project should be immediately applicable to similar simulations in theoretical astrophysics.

The distributed interactive execution and analysis of storm simulations is currently limited by disk and network speeds, as the simulation process output is in the range of 32 Mb/s to 320 Mb/s . The critical limitation today, however, is the conversion of data into graphic images, or visualization. Most three dimensional visualization today is done in batch mode using a mini- supercomputer which runs visualization software and can take between several seconds to several minutes to convert raw data into a single animation frame. This delay between simulation and graphic output prevents the researcher from interacting with the model and adjusting algorithms and parameters to yield optimal results. The delay also makes it impractical to collaborate with colleagues during the model verification process, as the scientist must send to the collaborator a finished product (a video) which will arrive several days or weeks after the simulation was done. Further, non-interactive visualization prevents colleagues from collaborating in the area of visualization techniques. Because each model is visualized differently, it is difficult at best to compare the validity of different models. Short term improvements in surface visualization will be obtained by using the supercomputer to do the tessellation component of the process (computing the geometric polygon representation) and to display the images in near-real time using graphic rendering hardware at a scientific workstation on the network. This will allow the researcher to interact with the simulation.

A sample collaborative session between two researchers at different locations could involve several components. Both researchers would have the capability of starting up the simulation or data analysis software from their workstation. Everything that appears on one of the researcher's workstation windows would appear on the other's (this requires screen transfers that can easily exceed 100 Mbits per sec for color). At any time either one of the researchers can take control of the process, or start up visualization from a different dataset for comparison. Surface displays from today's storm models consist of 30,000 to

100,000 polygons and can be coupled with other forms of visualization to qualitatively and quantitatively analyze model information. Animation of these displays being viewed at one site, also appear at the other. A high resolution animation of 8 bit per pixel images can be done with gigabit speeds (1400 frames would take about 15 s to transfer at 1 gigabit/sec). Data sets may have to be moved quickly from one researcher's site to another, depending on how the simulation and data exploration process is distributed and on the capabilities of the local graphics workstations. For collaborative interactive data exploration, this also requires gigabit transfer rates.

Long term improvements, only possible using a gigabit/second wide area network, will allow the display of simulation output , interactive control of the simulation and interactive analysis of the output to take place concurrently at multiple, separate workstations on the network. At this point, real-time collaboration will occur between scientists in the areas of modeling theory as well as visualization techniques. Each scientist will view the simulation variables of most interest to him and in a way which is consistent with the methods he uses to visualize his own model. Thus, the scientists can directly compare the output of two simulation models and begin to determine the strengths and weaknesses of the various modeling techniques.

Further development and increased workstation processing power will conceivably allow these scientists to do the tessellation as well as the rendering on their local workstation using their own custom visualization filters and to jointly analyze the simulation with colleagues across the network. Specific development will include a network software interface similar to the BSD sockets or to the Shared-X-Window system but with mechanisms for specifying experimental network services such as guaranteed minimum throughput, real time services, packet trains, isochronous data stream delivery, maximum tolerable latency, multi-cast, etc. to be implemented by network researchers on the testbed. Further investigation will be made into the transmission over the network of multiple channels to provide voice and image teleconferencing in parallel to simulation output and control.