National Radio Astronomy Observatory

COMPUTER ADVISORY GROUP DOCUMENTATION

AN NRAO PLAN FOR A LARGE COMPUTER

August 1984

Table of Contents

ABSTRACT

In recent years radio astronomers have made increasingly heavy use of synthetic aperture telescopes, such as the VLA, to obtain ever higher resolution and sensitivity. Since the optics of such instruments are formed by computers, major improvements can be made to the instrument well after the hardware is constructed. This flexibility has led to dramatic improvements in both the quality and quantity of the images produced by the VLA which now greatly exceeds its original goals.

This enhancement of the VLA has come at the cost of greatly increased computer usage but the available computer facility has not kept pace. At this time only a small fraction of the very computer intensive, but scientifically interesting, investigations are actually being carried out.

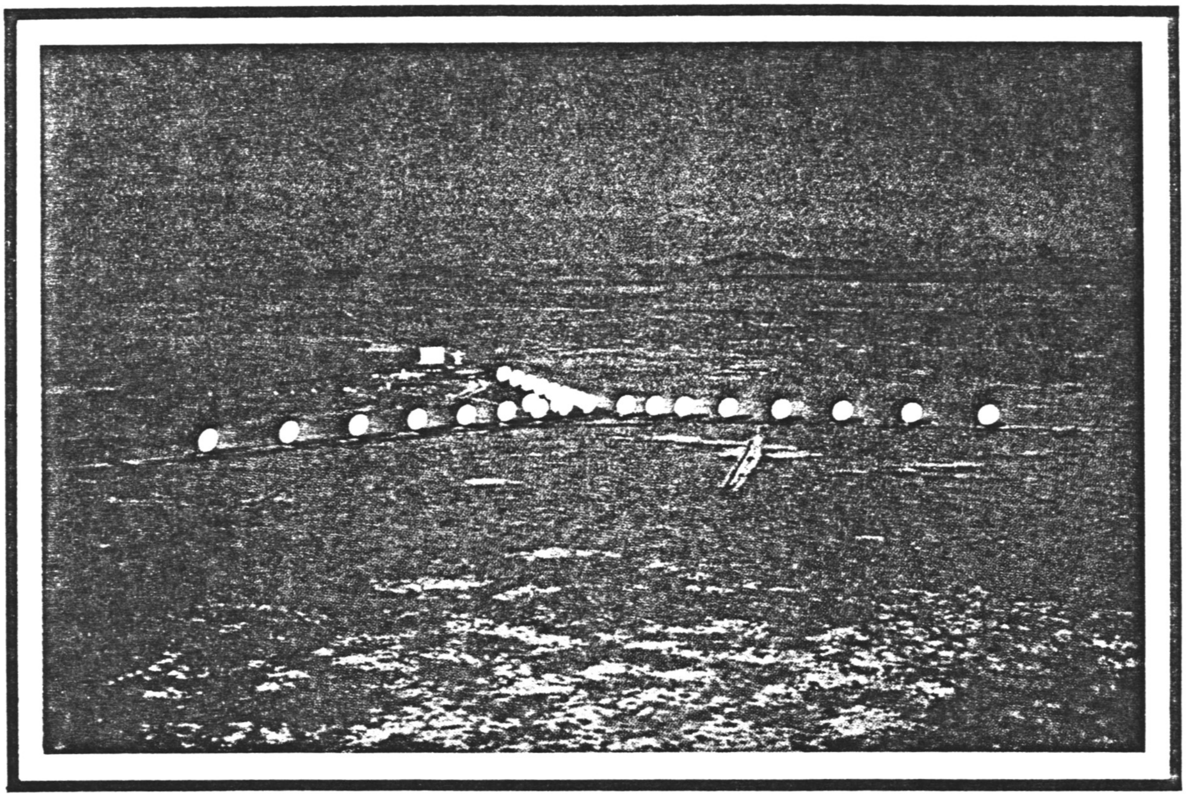
In order to specify the computer capacity that will be necessary for image formation and enhancement, NRAO has estimated the demand for VLA observing and consequent data processing over the next several years. Similar estimates have been made for the VLBA and for other anticipated projects. From this work, it is clear that computer demands will exceed the present NRAO capacity by at least a factor of 10.

NRAO computer needs, calculated from the anticipated astronomical requirements, call for access to an achieved computer capacity of 60 to 80 megaflops. This power, in a system with sufficient memory and I/O capacity to avoid serious bottlenecks, can be provided by the equivalent of about 100 VAX-class machines or 1 relatively large supercomputer, i.e. a class VI computer (e.g. a Cray XMP).

Over the past several years NRAO has invested a substantial effort to making a major portion of its software (the AIPS package) truly transportable. Thanks to this prior effort, NRAO is now in a position to make effective use of a supercomputer almost immediately.

Although various options for sharing a computer facility might seem to be attractive as ways of obtaining the extra computer power, a closer look at each option leads to the conclusion that NRAO should continue to own and manage its own facility.

The total capital cost for such a facility is estimated to be $16.9 million, with an annual operating cost of $5.3 million which is $2.4 million above NRAO's current computer budget.



Aerial view of the Very Large Array
as seen from the northeast

1. - INTRODUCTION

(A brief background overview to the process of Aperture Synthesis including data sampling, calibration, image formation, and image processing is to be found in Section 2 of the Documentation for the Computer Advisory Group meeting in Socorro, March 1982.)

Synthesis radio telescopes can now make detailed images of the radio sky with better resolution than optical telescopes. A conventional monolithic imaging telescope with equivalent resolution would require an aperture of many kilometers. It would be prohibitively expensive and probably impossible to construct. The digital computer has made possible the synthetic radio telescope, which forms images from measurements of the coherence of the wavefront, and now forms an essential component of such a telescope.

The National Radio Astronomy Observatory (NRAO) has one major synthesis telescope in operation, the Very Large Array (VLA), and another, the Very Long Baseline Array (VLBA), is being funded in 1985. The VLA is the largest and most sophisticated radio astronomical facility in the world. Construction started in 1973 and the completed instrument has been operating since 1980 (Figure 1). Since the original design of the VLA, many enhancements to its capability have become possible:

i) the capability to form images simultaneously in many frequency bands was added, enabling determination of the velocity, as well as the spatial structure, of radio emitting objects.

ii) in the original concept, it was assumed that an image would take from 8 to 12 hours to build up, but further development of image deconvolution techniques allowed vast reductions in the observing time so that now hundreds of images can be made in the same period of time.

iii) astronomical observations made with the VLA have shown that the radio sky contains vastly more small scale structure than had been anticipated. As a result, the typical size of the images has increased by factors of 4 to 8 in each direction.

The net result of all these enhancements is a peak output data rate almost a million times greater than that anticipated when the telescope was first proposed. Although the capacity of digital computers also increased greatly during this period, they did not increase by the same factor, so the science which can be performed with the VLA is now limited by the computational resources available. In addition to the increased load caused by the enhanced VLA capacity there have been two revolutionary developments in image processing which have increased the computer requirements by a further order of magnitude.

In order to explain the importance of the new developments, we first have to describe the prior major limitations of synthetic arrays. Images formed by synthetic arrays, such as the VLA, are corrupted by two effects: i) incomplete measurement of the coherence of the wavefront due to the finite number of elements and limited observing time, and ii) wavefront collimation errors due to the spatially and temporally varying refractivity of the atmosphere above the array.

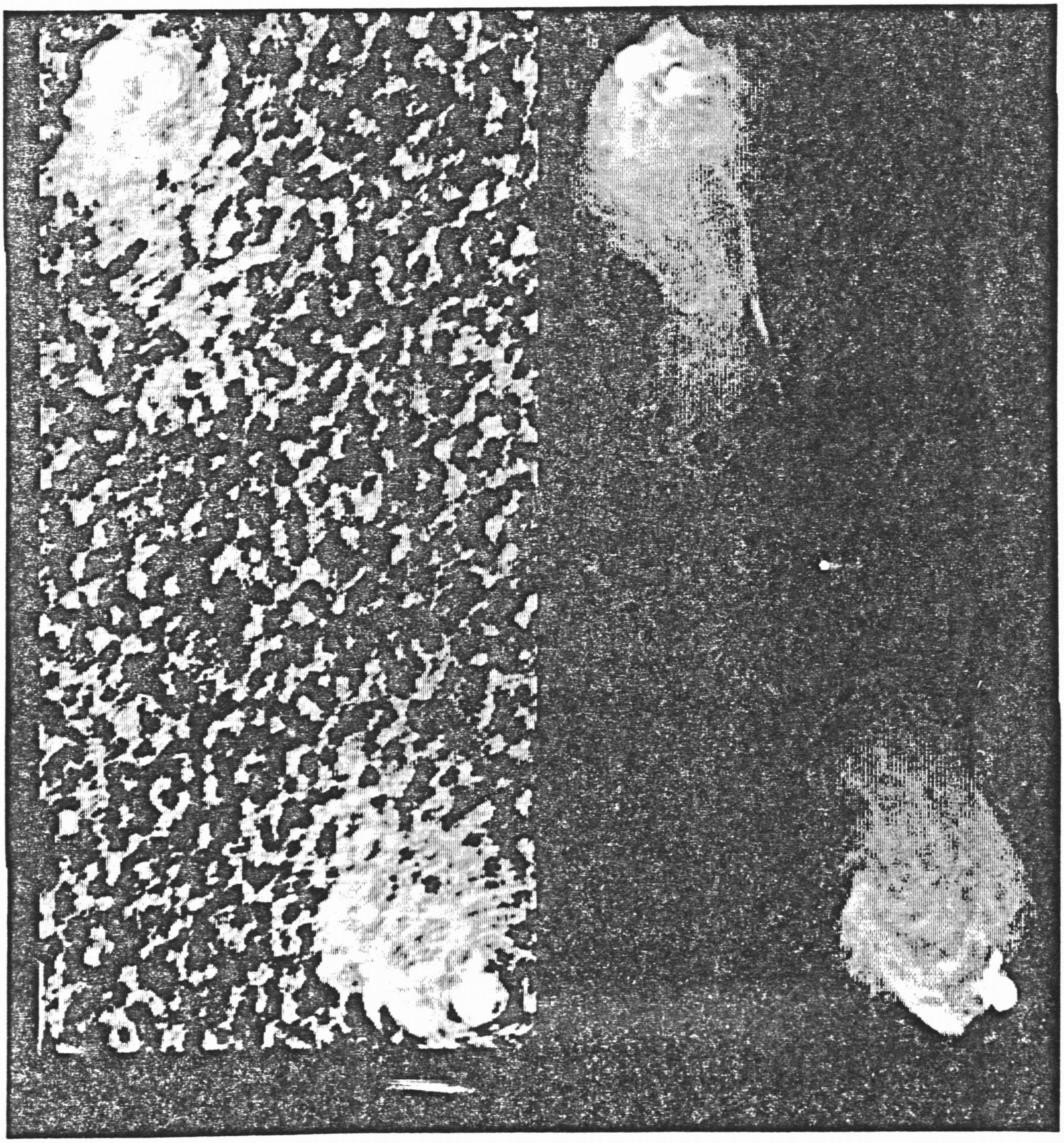
The first effect has no analog in optical telescopes since optical images are formed by a continuous lens. In radio synthetic imaging the missing information can be filled in if one can make some simple, and usually valid, assumptions. For example, if the object to be imaged is of limited extent then the coherence of the wavefront must change slowly. Unsampled measurements can then be interpolated and an unblemished image formed. Such assumptions form the basis of new deconvolution algorithms which have been developed within the last ten years.

The second effect, imperfect collimation, is a major obstacle to ground based optical imaging since, for most types of observations, it limits the effective resolution achievable to about 1 arcsecond. In principle, the collimation can be adjusted in real time by the use of a "rubber mirror" in the optical system to cancel out the collimation errors. In practice, however, the rapid twinkling of the atmosphere at optical wavelengths limits this technique to observations of only the very brightest objects. Furthermore, the mechanical problems of flexing a mirror to tolerances of an optical wavelength at rates up to 100Hz are formidable. In radio synthetic arrays an essentially identical technique which has been developed over the past 8 years by various groups (selfcalibration) is very much more useful since, firstly, the twinkling time of the atmosphere at radio wavelengths is measured in minutes rather than milliseconds and secondly, the collimation corrections can be applied off-line to the digitally recorded measurements of the wavefront coherence.

Used together, these two techniques, deconvolution and selfcalibration, have increased the dynamic range of the best VLA images from the 100:1 specified in the original proposal to over 10,000:1, but at the cost of greatly increased computation time. The impact of selfcalibration has been even greater in Very Long Baseline interferometry where true imaging of objects at resolutions up to a thousand times greater than the VLA's capability is now possible.

As an illustration of the increased imaging power of the VLA using these techniques we show in Figure 2 a "before and after" comparison of the VLA images of the archetypal radio galaxy Cygnus A. The "before" image obeys the design specifications of the VLA but portrays far less detail than the "after" picture, and, in particular, it shows neither filaments in the extended structure nor the thin line of emission connecting the central core, which is the nucleus of the parent galaxy, to the right lobe. The wealth of unexpected detail revealed in the "after" image dramatically enhances its scientific importance.

We must, of course, accept that there will always be factors which limit the capability of any instrument. The receivers and antenna area determine the faintest objects which can be observed and the size of an array sets limits on the highest angular resolution which can be



(a) (b)

Fig. 2. The powerful radio galaxy Cygnus A.
a. Before computer image enhancements, and
b. After use of the self-calibration
and deconvolution algorithms.

achieved - such hardware limitations are well known and accepted. The limitations to productivity caused by the finite computer resources occur in a slightly less direct way. By spending inordinate amounts of time using small computers, or networks of small computers, any single project may still be possible. The limitation imposed by the computer is one of total throughput and it may be especially insidious because its effect is indirect.

As remarked in the Curtis report to the National Science Foundation (NSF), the unavailability of supercomputers has inhibited activity in certain scientific fields. This situation certainly occurs with synthesis radio telescopes: the best astronomers, conscious of this limitation (whether real or perceived) will simply turn to other problems and the most imaginative and the most important problems in radio astronomy may not be done. For observations which are just feasible with the present resources the long turnaround time inhibits both experimentation with analysis methods and investigation of unusual effects. The process of scientific discovery is a fragile one. Limited computer resources which increase the barrier between the scientist and his observations interfere with this process and will not be conducive to doing good science.

The thrust of this long term plan is to address the imbalance between the computer resources available and the investment in the rest of the telescope. The construction cost of the VLA, scaled to 1984 dollars, is $130 million. The proposed VLBA is an additional $60 million. The cost of the supercomputer proposed in the accompanying documents is less than 10% of this capital investment yet it is likely to improve scientific productivity by a large factor.

VLA Scientific Memorandum No. 150 ("Astronomical Requirements for Future VLA Processing") analyzed the projected scientific requirements in terms of their data processing load to make a detailed estimate of the actual computing capacity required for the VLA. This analysis consisted of the following steps:

i) Listing the astronomical projects which will use the VLA and tabulating the relevant observational parameters.

ii) Determining the required processing capacity on the assumption that the astronomical objectives should not be compromised by the computer, while accepting any other hardware limitations.

iii) Estimating the fraction of observing time that would be spent on each project based on a projection of the current use of the VLA. This estimate is necessary to calculate the time averaged throughput required given the widely varying amount of processing required for different projects.

In order to convert these astronomical requirements, which are specified in terms of parameters such as image size, into computer hardware requirements, VLA Computer Memorandum No. 168 ("A Computer Plan for the VLA") analyzed the imaging process in terms of the number of floating point operations and the I/O bandwidth required. Their analysis involved a hypothetical computer of conventional architecture, with sufficient main memory to avoid using disk I/O for sorting, and with computational power of five million floating point operations per second (5 Mflops). The results of this analysis have been scaled to other computers by the ratio of the Mflops ratings. Figure 3 summarizes the results of these two investigations.

In Figure 3 the various categories of astronomical projects are indicated by blocks in which the abscissa is the computing time in hours required for each day of observing. The time scale is set for a Class VI supercomputer with 60 to 80 Mflops continuous achieved capacity. For comparison with our current computers, the time scale applies to a VAX-11/780 with an AP-120B array processor if the units are days instead of hours! The ordinate is the fraction of the observing time which would be devoted to each class of observation. Thus the time averaged throughput corresponds to an area in this diagram and the capability, in terms of turnaround time for a given observation, is given by the extent of each block in the horizontal direction. The shaded area indicates the total capacity of the proposed 60 Mflops (sustained average) supercomputer, and the dotted inset shows the total capacity of all NRAO's present computer systems.

Two conclusions can be drawn from this diagram: firstly, the necessary throughput is sufficient to require constant use of a 60 Mflops computer and secondly, such a system would be capable of supporting analysis of the majority of the proposed observations in a time comparable to the observing time. Both of these aspects are requirements of the future system.

As emphasized in the Curtis report, the capability to perform a given computation in reasonable time is also important for scientific productivity. This requirement excludes a system assembled from many small computers, which might still provide the required throughput but not the capability in terms of the turn around time for big projects. To support this argument, we have redrawn in Figure 4 the data from VLA Scientific Memorandum No. 150 and VLA Computer Memorandum No. 168 as a cumulative distribution. This shows that 40% of the astronomical projects can be handled in less than real time with our present VAX and array processor (AP) combination, and explains why much good science is being done even with the limited computing resources now available. This kind of project is ideally suited to the superminicomputer systems both at NRAO and in universities. However, the diagram also shows that 30% of the projects would require more than 30 days CPU time with a system this size, a figure which our experience suggests would be totally impractical. The line scaled for the 60 Mflops supercomputer shows that 80% of projects can be computed in less than real time and that 95% can be handled with a turn around not much longer than real time.

To make these points more concrete it is useful to look at some specific examples based on actual projects which have been attempted with the currently available resources. The mapping of the remnant of the supernova explosion in Cassiopeia A is currently in progress. The scientific objective is to determine the distribution of intensity and polarization of the radio emission with a resolution of 0.7" over the

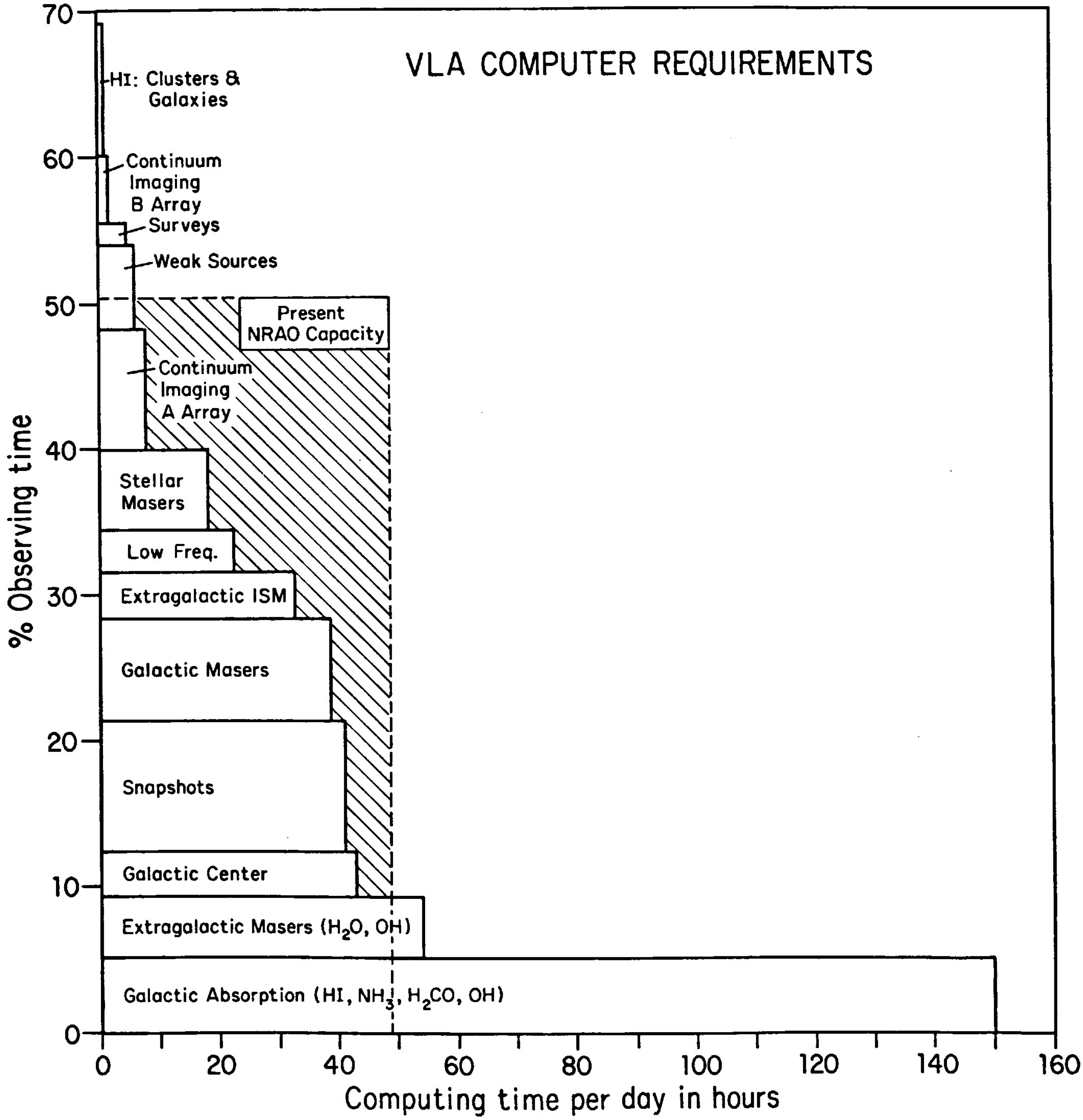


Fig. 3. Computer resources required for various categories of astronomical projects. The horizontal extent is the computing time in hours required for each day of observing assuming a 60 megaflops supercomputer. The time-averaged throughput corresponds to an area in this diagram, and the turn-around time for a given observation to the horizontal extent of each block. The shaded area indicates the total capacity of the proposed supercomputer while the insert indicates the total current capacity of all NRAO computers.

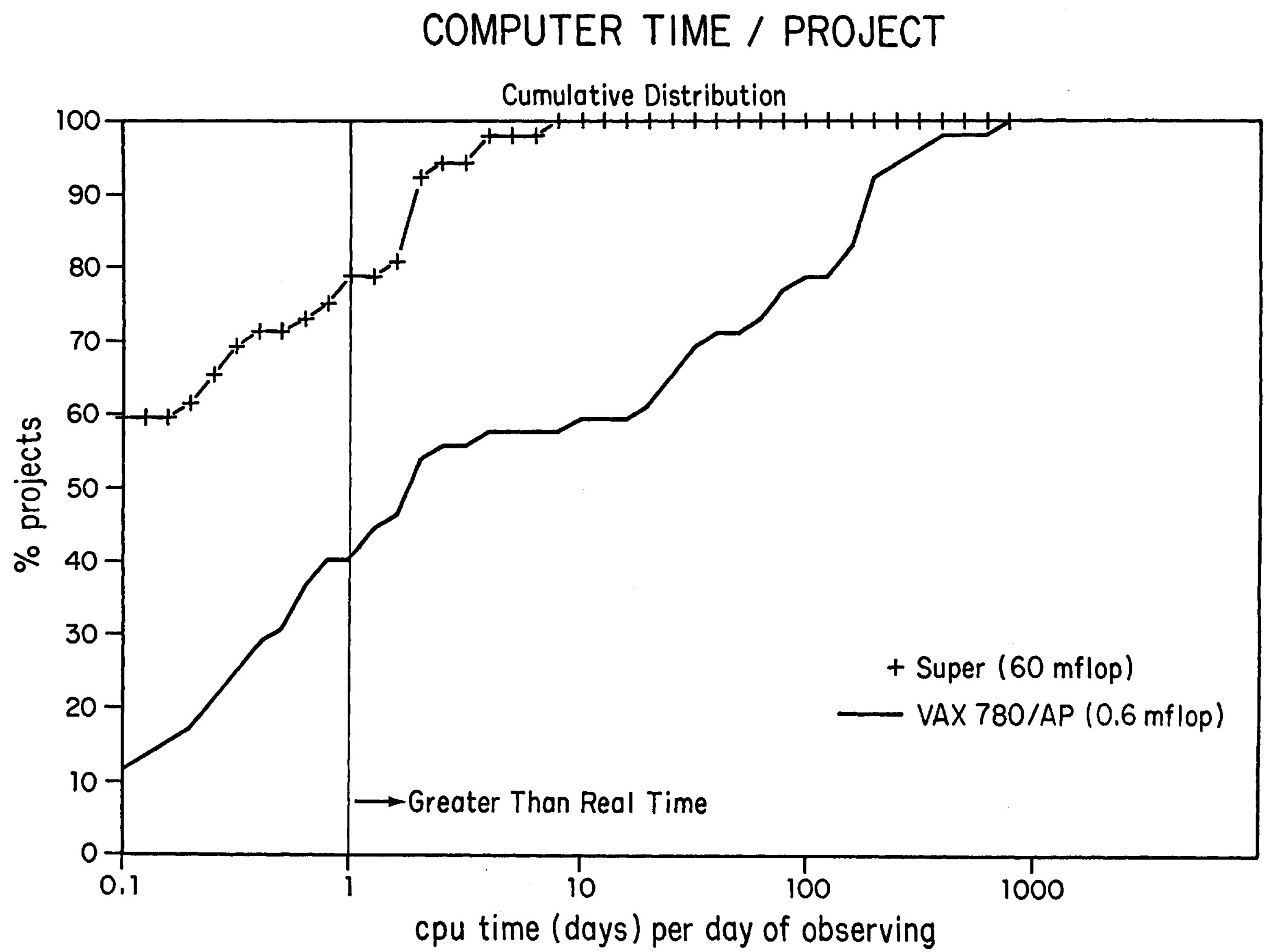

COMPUTER TIME / PROJECT
Cumulative Distribution
% projects
100
90
80
70
60
50
40
30
20
10
0
0.1
1
10
100
1000
cpu time (days) per day of observing
+ Super (60 mflop)
VAX 780/AP (0.6 mflop)
Greater Than Real Time

Fig. 4

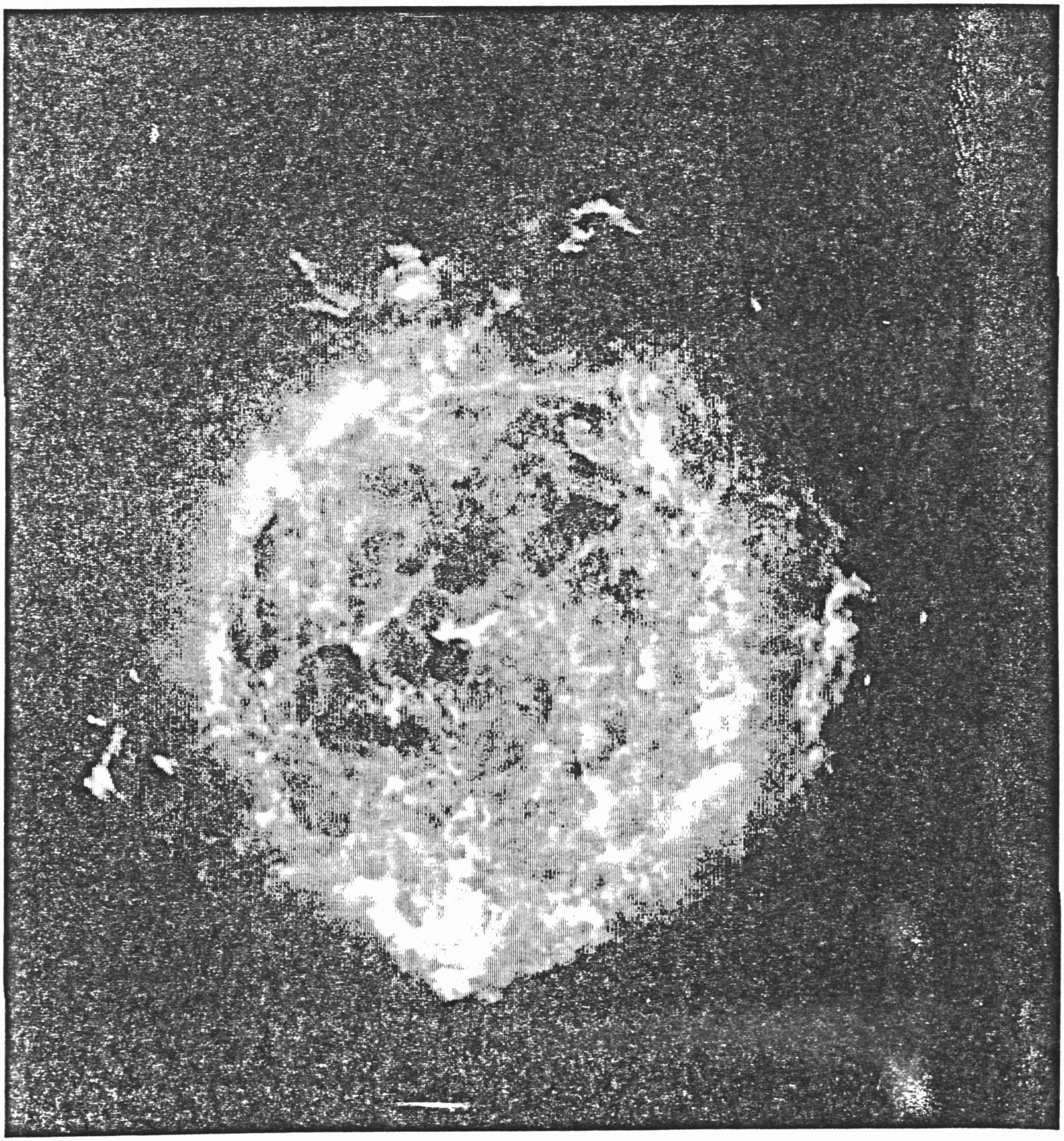


Fig 5. The remnant of an exploding star in the constellation of Cassiopeia. This the result of a deconvolution of a 2K x 2K image. It required 3 days computing time on a VAX 11/780 + AP120B but has only 1/3 the possible resolution.

whole remnant of size 6'. This will be compared with the existing infrared, optical and X-ray images, and used to study the developments of turbulence, and, eventually, to measure the motion of the turbulent features. The data result from a total of 26 hours observing, and comprise 5 million independent measurements of the wavefront coherence stored on 20 1600 bpi magnetic tapes. Conversion to low resolution (2k x 2k) images (Figure 5) used three days of computer time on an empty VAX-11/780 with an AP-120B. The desired full resolution (4k x 4k) map will take 12 days computing. Final correction of this image would require a further few months of computer time. Projects of this scale could be processed in less than the observing time on the proposed supercomputer.

The second example involves the determination of the mass distribution in the central region of our galaxy from the distribution of the velocities and the positions of a special type of star which emits a sharp radio frequency line due to OH maser emission in its atmosphere. In this type of observation the OH maser stars are found by searching in a three dimensional (two spatial and one velocity) cube. While the observing time for this project is 5 days, the resulting images, which, with the present array hardware, consist of 12 1k x 1k x 512 cubes, can only be computed in 10 days using our present VAX and AP system. In a few years, hardware changes to the array will make it possible to increase the depth of the cube to 1k pixels, and to obtain the cubes in each of two orthogonal polarizations simultaneously for two of the OH lines. The same observing time will then require 110 days computing time in the VAX and AP system! The final result of this experiment is a list of the positions and the velocities of the few hundred stars which will be found in these cubes. Hence the final analysis leading to the mass distribution in the center of our galaxy is a relatively simple step. In projects of this type, the extensive computations are only required as an intermediate step, and the final result involves analysis of a very modest amount of data.

The preceeding discussion and examples were all based on experience with the VLA. Although there are other synthesis telescopes either now being funded, such as the VLBA, or which will be proposed in future, such as the millimeter wavelength array and the orbiting satellite arrays, we cannot make such detailed analysis of their requirements before we have any observational experience. Crude estimates indicate that these future telescopes will have requirements probably less than, but perhaps comparable, to that of the VLA.

In summary, we argue that it is a necessary and an appropriate role for NRAO to provide computer hardware facilities of this magnitude because such facilities are not readily available to the majority of our users. Since the host institutes for our user community are already becoming well equipped with computers in the superminicomputer class, it would be appropriate for NRAO to concentrate on acquisition of hardware with a very much greater capacity in order to provide the opportunity for all users to do the more difficult observations. At present, only those privileged users who have access to superminicomputer systems for very long periods can do the large projects. NRAO should provide sufficient computer capacity for any user of the telescopes to obtain images in a reasonable time.

2. - THE CURRENT STATUS OF NRAO COMPUTING

The present NRAO computing facility consists of some thirty relatively small machines buttressed by ten array processors. A summary of this equipment is shown in Table 1. The last column in this table is an estimate of the computing power that can be realized from each item in Mflops because the vast majority of NRAO computations involve floating point operations.

Table 1

Equipment List

System	Computers	Array Processors	Estimated Mflops
VLA On-line system	8 ModComp-II	1	3.0
VLA Asynchronous system	DEC-10		0.25
VLA Pipeline	4 PDP-11	4	6.0
VLA AIPS	2 VAX-11/780	2	3.0
Charlottesville AIPS	VAX-11/780	1	1.5
	ModComp-IV	1	1.5
Charlottesville General Use	IBM 4341		0.25
Charlottesville VLBI	2 (small minis)		0.05
Tucson General Use	2 VAX-11/750		0.25
Tucson On-line system	3 PDP-11		0.5
Green Bank On-line system	6 (various)		1.0
Green Bank General Use	Masscomp		0.25
	ModComp-II		0.25
Totals	33 machines	9 APs	18 Mflops

The aggregate of hardware currently in use at NRAO represents a substantial resource, equivalent, at least at face value, to several of the largest conventional computers.

It is clear from Table 1 that APs represent a very significant portion of all the presently available computer power at the VLA. The APs dominate even though the estimated obtainable power given in Table 1 is only a fraction of their rated maximum performance (12 Mflops per AP). Unfortunately, even the fraction quoted cannot be used effectively.

The major reason for this problem is that in most NRAO machines, and especially in its APs, processor speeds are not matched by memory size nor by I/O (channel) capacity. In other words, memory and I/O are severe bottlenecks. What is worse, the bottlenecks are inherent in the equipment design so that they cannot be relieved by the simple addition of components.

The magnitude of this mismatch can be appreciated by noting that the APs in the Pipeline are meant to produce maps of 2k x 2k pixels routinely and maps of 8k x 8k pixels occasionally. Yet the largest AP in the Pipeline contains only three pages of memory (1 page = 64 kwords) which is less than 2% of the space required for a single routine map. Since the gridding machine cannot hold a full map the input data must be sorted (at least partially into bins). This makes things awkward when the natural time-baseline order would be appropriate, e.g. when editing or calibrating the data.

Even the PDP-11s, which contain relatively large amounts of memory, are afflicted with bottlenecks because the operating system restricts each job to 64 kbytes of usable address space. (A much larger memory space can be used but only for data and only at a great loss of efficiency incurred from register swapping.)

The other condition that limits the available computer power might be called "computational entropy" by analogy to the factor that limits the amount of work derivable from thermal energy. In computing, as in thermodynamics, diffuse power is less usable than concentrated power when that power is applied to problems that cannot be partitioned.

Consider the VLA on-line system, for example. Even apart from the equipment allocated to working spares, some of the ModComp CPUs are idle while others are near saturation. The spare cycles cannot be used to relieve the overburdened machines because the CPU-limited tasks cannot be shared. The on-line system uses only a quarter of its AP because the amount of data to be processed is limited by the rest of the on-line system.

The Pipeline, which has recently come into production, suffers severely from similar computer entropy problems. At least half of the Pipeline capability is wasted shuffling data from one machine to another via shared disk and converting the data from one format to another. For a further discussion of this problem the reader is referred to VLA Computer Memorandum No. 172 ("Whatever happened to the Pipeline?").

3. - TECHNICAL OPTIONS

There is no single conventional computer that can provide the computational power required by NRAO (60 to 80 Mflops continuous effective rate). The only means by which such large amounts of computing can be obtained are:

i) A special purpose machine.

ii) A multiplicity of machines, each supplemented by special peripherals (Array Processors).

iii) A Supercomputer.

3.1 - Special Machines

Judging from the number of published articles on work in progress there is a considerable amount of research underway on alternate computer architectures. In almost all cases, the articles report on efforts to create a very powerful computer from a large number of relatively inexpensive chips, exploiting parallelism and pipelining in fundamentally new ways. The development of systolic machines is typical of this work.

It is possible that, within a short time, current research will lead to a new generation of supercomputers. For now, however, the only hardware that has come from this research must be classified as experimental rather than pre-production. Anyone planning to use such machines must anticipate a lack of software (including appropriate high level languages), maintenance, and vendor support of any sort.

There is a single exception to the situation just described. Denelcor's Heterogeneous Element Processor (HEP) is a commercial product which is supported by the vendor in all the traditional ways. The HEP however has not yet been delivered in a configuration that can supply anywhere near the number of compute cycles required by NRAO. And while the HEP is undeniably a "production" machine, the techniques required to program any MIMD device are still being developed. In fact, existing HEP machines seem to be used primarily to support research in this field.

Since NRAO is not in a position to participate in the development of experimental computers, (the Computer Advisory Group itself strongly recommended against experimental computer activities), new special purpose machines must be ruled out of current plans.

3.2 - Multiple Machines with Array Processors

Even though no single mainframe machine can provide, by itself, the level of computing power required by NRAO, a set of conventional machines buttressed by APs could, at least in principle, supply the necessary resources.

NRAO has had extensive experience with this approach. Over the past decade, NRAO has tried to acquire high computing power at low cost by interconnecting multiple devices in a configuration tailored to its image processing requirements. The "Pipeline," a network of several minicomputers, four APs, special memory hardware, numerous special (dual-ported) disks and tapes, has just recently been placed into production.

One of the most valuable outcomes of the Pipeline project is the lesson that has been learned concerning the need for flexibility and the very high price that is inevitably paid for complexity. For further elaboration on this point the reader is referred to VLA Computer Memorandum No. 172 ("Whatever Happened to the Pipeline?").

From the lessons learned during development of the Pipeline and from the difficulties currently being experienced keeping a multi-component distributed system in operation it is painfully clear that the value of directness and simplicity in a single hardware and software system greatly outweigh any advantage that can be gained by a clever combination of less expensive components.

The current AIPS computers also depend heavily on APs but in a simpler configuration than the Pipeline. AIPS machines consist of a single general purpose computer with an attached AP. This configuration has allowed AIPS to design and implement new data processing algorithms, e.g. selfcalibration, which would be exceedingly difficult on the less flexible Pipeline.

Both the Pipeline and AIPS implementations of APs depend heavily on the use of custom microcoded software to use the power of the AP. Microcoded software is very difficult to write and extremely difficult to debug but is absolutely vital for the efficient use of APs. This is likely to remain true for the forseeable future.

The difficulty of writing microcoded software seriously retards the process of developing new data processing techniques. Many of the operations required for processing synthesis telescope data are sufficiently unique that commercial software is not currently adequate nor is it likely to become so in the future. This means that the development of new, high performance software for APs is limited to a few persons trained in writing efficient microcode. It virtually prohibits a large pool of scientists from trying out new ideas.

Thus, while APs appear capable of delivering the needed computing capabilities, their inflexibility seriously inhibits future development of data processing techniques. Since this development has, in the past, greatly increased the scientific output of the VLA we feel that future dependence on APs is an undesirable impediment to the development of the VLA and synthesis telescopes in general.

3.3 - A Supercomputer

The term "supercomputer" refers to machines made by only two American manufacturers. Cray Research offers several models; Control Data Corporation (CDC) offers its Cyber 205. New models will soon be available from Cray and the TF10 is being developed by ETA Systems, a CDC spin-off company. Among Japanese manufacturers, Fujitsu and Hitachi either have announced supercomputers or are expected to do so in the immediate future. Some of their machines may become available through an American partner or subsidiary.

Supercomputers are distinguished from other large but conventional machines by their logical organization (architecture) which departs from the von Neumann model followed by all other computers since the ENIAC. For the purpose of this discussion, the defining feature of a supercomputer is the presence of one or more pipelined arithmetic units which are used to implement vector instructions.

Since the hardware is pipelined, supercomputers can initiate a new operation each cycle even if it takes several cycles to complete; several arithmetic units may run in parallel. This combination of pipelining and parallel processors produces a considerable increase in computing speed but is only practical when the same operation is to be done to a relatively large number of values. Since different supercomputers have different vector operation startup costs, the length of the typical vector operation to be performed influences the choice of supercomputer.

Cray is developing a disk I/O capability (striping) in which a file is scattered onto several disk drives with independent controllers and data is multiplexed onto different drives. With this scheme, the effective I/O bandwidth for a given process can be multiplied by the number of disk drives. This increase in I/O performance will improve the match between computation and I/O for synthesis array data processing.

Unfortunately, there is no characteristic vector length in processing synthesis data; vector lengths range from thousands (big FFTs) to a few (7 in u,v data gridding). In principle, with suitable modifications of algorithms, a problem involving short vectors can be stated in terms of longer vectors. In the case of gridding however, this restatement can lead to data dependency.

In pipelined operations there is a fundamental assumption that all input and output data values are independent. In the task mentioned above, gridding data observed at quasi-random points, each observed value is convolved with a function whose support size is typically 7 x 7 cells; results are then summed into a grid. In general, several observed values will contribute to a given cell. The net effect is that, when the gridding task is formulated as a long vector operation, the input value of one operation is frequently the result of a previous step in the same vector operation. That is, the input values are not independent. This problem is especially serious on CDC 205 machines.

3.3.1.1 - Evaluating Cray Machines for NRAO Applications

The vector processor in Cray machines is relatively well suited to short vectors; in fact, vector operations involving more elements than the length of vector registers (64) are automatically broken into steps of that length. Much of the power of the Cray vector processor could be utilized by NRAO gridding routines in which the inner loop is typically small. Vector operations in a Cray can have an arbitrary but constant stride (interval between the elements of a vector).

The biggest disadvantage of Cray architecture is the 64-bit word size. That is double the size needed in NRAO applications. (A dynamic range of 10,000:1, the present VLA limit, requires no more than 14 bits and a few guard bits.) Although it may be practical to store at least some data and intermediate results in a packed format (two values per word) in order to reduce storage and I/O bandwidth requirements, the unnecessarily long word remains wasteful because the hardware does not support any 32-bit operations.

The main memory on Cray machines is quite expensive and comes in relatively limited amounts compared to the memory sizes currently available on conventional mainframe machines. The limited memory would inhibit such typical NRAO operations as gridding unsorted data. Cray has a solid state disk drive (SSD) with up the 32 Mwords of memory. This fast bulk storage device can be used to alleviate some of the problems that arise from small main memory size. The Cray-2 will have a virtual memory system in addition to much larger physical memory so that the memory size will not be a problem (except for the cost!).

The Cray-1 is rated at nearly double the Mflops value that needs to be achieved on a continuous basis to satisfy the stated NRAO requirement - 165 Mflops rating versus a requirement of 60 to 80 Mflops sustained average. But those who have been using Cray-1 computers the longest, the Los Alamos National Laboratories (LANL) and the National Center for Atmospheric Research (NCAR), report that their sustained rates for a Cray-1 are about 30 Mflops or only half of what NRAO needs.

To achieve the desired computing power from a Cray machine, NRAO would have to use at least a single processor model XMP machine which is rated at 220 Mflops sustained for vector operations, a sustained scalar rate of 125 million instructions per second (Mips) and 60 million combined arithmetical-logical operations per second (Mops).

3.3.1.2 - Evaluating the CDC 205

The biggest technical advantage of the CDC 205 for our application is that 32-bit floating point operations are supported. Furthermore, the 32-bit operations run at twice the speed of 64-bit operations. This eliminates the need to pack and unpack 32-bit data.

Other advantages of the CDC 205 are the availability of more physical memory than on a Cray and the availability of a virtual operating system to exploit that memory. It is not clear if CDC will support a semiconductor bulk storage device similar to the Cray SSD for the 205.

The speed of a CDC 205 can be increased by adding more arithmetic pipes with no change in the software. Multiple pipes are always used in parallel.

The biggest disadvantage of the CDC 205 is that its vector operations obtain a reasonable fraction of the theoretical speed only for long vectors (about 50 elements minimum for a single pipe machine doing 64-bit operations). The minimum efficient vector length increases as the number of pipes in the CPU (one to four) and doubles for 32-bit operations. The principal reason for this is the large startup overhead of 50 cycles (one microsecond) for vector operations.

In the CDC 205, vector operations can have only unit stride. This means that, if the elements of a vector are not already in contiguous memory locations, they must be collected into a temporary contiguous vector before the vector hardware can be used and the results must be dispersed subsequently. This requirement is particularly onerous when the natural form of the data is complex (as in the case of aperture synthesis data) since real and imaginary values are normally interlaced. The 205 does have fairly efficient, hardware gather/scatter operation which could mitigate this problem.

3.3.1.3 - A General Comment on Supercomputer Hardware

The configuration used for all CDC 205 machines and for all Cray machines running COS (the Cray Operating System) includes one or more front-end computers. This is the machine with which users communicate and through which users submit batch jobs to the back-end supercomputer. Users have no direct contact or control over the supercomputer itself. It is also common on batch supercomputer systems to have only a small amount of fast scratch disk space; such a scheme requires transferring all input data and all results between the machines for each execution of each task. There are also problems of interprocess communication between the two machines.

Since synthesis array processing frequently requires large amounts of input data and large output images, the files to be transferred between the front-end and the back-end machines frequently would be very large. What is worse, in most front-end/back-end configurations the two machines have different word sizes, different word formats and different file structures. Translation is required in addition to transportation. So the load imposed by such traffic rests on the CPU as well as on I/O channels.

Clearly the usual front-end/back-end configuration has inherent problems that could strangle any supercomputer intended for NRAO application.

3.3.2.1 - Software Options (General)

The use of a supercomputer for synthesis array data reduction must satisfy two conflicting considerations: high efficiency and quick response. The requirement for high efficiency means that the machine should run jobs in at least a quasi-batch mode. In this fashion, the resources of the machine can best be scheduled and managed. Unfortunately, batch systems are notoriously unresponsive. A traditional batch system will greatly reduce the productivity of the scientists using the system and reduce the scientific output of the system.

A compromise between a traditional batch system and a single user interactive system is a multi-user interactive system with a background batch stream. This allows better management of the system resources while retaining an interactive user interface. In most cases a given user would get only a fairly small fraction of the total machine resources. This means that the real time for processing a given set of data will take considerably longer than for a single user supercomputer.

A further consideration for supercomputers has been briefly described above; namely, the front-end/back-end, batch oriented nature of many supercomputer installations. In these configurations the front-end machine essentially prepares a virtual card deck to feed to a virtual card reader of the back-end machine. This configuration works reasonably well when there is little data but a great deal of computation involved. This is not our case.

For a number of reasons, it is desirable to have multiple users talking to a single machine which has all of the disks, display devices, tape drives etc. This configuration would allow relatively efficient use of the hardware while maintaining a user friendly working environment.

3.3.2.2 - Specific Software choices

The CDC operating system for the Cyber 205, VSOS, is capable of supporting interactive use. CDC claims that all or almost all installations run 205s in a batch mode with a front-end machine. It is not clear if there is a fundamental problem with using VSOS for interactive use or all users of 205s have batch oriented problems.

There are currently two operating systems available for a Cray; Cray's own system (COS) and an interactive system (CTSS) developed by Livermore Laboratories. COS is a front-end/back-end batch processing system which is not well suited to our kind of data processing environment.

CTSS allows interactive use of the Cray directly and is thus much closer to the environment we currently enjoy. Cray Research is developing a Unix operating system for the Cray-2 and may release it for the Cray-1. A Unix system, if NRAO could get it, is probably the most ideal.

3.4 - Conclusion

All things considered, the current system of choice is a single processor Cray XMP with a substantial amount of SSD running under CTSS (or Unix, if it becomes available). The principle reasons for this choice of hardware are i) good performance for short vectors, ii) the prospect of good I/O performance, and iii) the availability of an appropriate operating system.

4. - MANAGERIAL OPTIONS

While it is clear from the foregoing that NRAO needs a relatively large increase in computing power, nothing that has been mentioned so far dictates the direct whether NRAO should operate its own dedicated system or subscribe to a shared facility.

There are four possible scenarios in which NRAO might share a large computer facility:

i) NRAO runs a large computer facility but, at the request of the astronomical community, other astronomical data processing is supported for purposes unrelated to NRAO activities;

ii) Another national facility, such as the National Optical Astronomy Observatory (NOAO) or NCAR, runs a large computer facility with some portion available to NRAO users and staff for NRAO-related data processing;

iii) Part of a large computer facility funded by NSF at a university is provided for use by NRAO, and NRAO users; and

iv) NRAO makes used of a large computing facility which is not funded by NSF, such as one of the supercomputer systems at LANL, Sandia Laboratories, etc.

4.1 - General Comments on Sharing

The principal disadvantage of all four of these options is that they are likely to be restrictive. NRAO needs are sufficient to require, by themselves, all of a supercomputer's capacity. It is very unlikely that, in any of the scenàrios listed above, the supercomputer would be large enough to handle all of NRAO's needs and have enough extra capacity to accommodate the other users.

The second important disadvantage, common to all the options mentioned above, is that each would involve the selection of a configuration that may be far from optimum for our application. Not only the hardware configuration, but also the choice of operating system, and the tuning of the operating system would be much different for a dominantly theoretical applications. It may well involve expensive additions to provide the support for both types of problems.

Thirdly, there is the specter of a disproportionately large administrative overhead. It will be enough of a management challenge for NRAO to provide supercomputer hardware, software, and administrative support for its traditional radio astronomy community of users without adding the classical problems that arise in any landlord-tenant relationship and without having to deal with a community of users or administrators totally alien to NRAO's goals and expertise.

To use another organization's facility, NRAO would have to provide considerable on-site support. This would lead to a further fragmentation of its already split programming and support staff. NRAO users have also come to expect a personal support from NRAO astronomers, observing "friends," and data analysts in addition to the usual computer operations support. It would not be possible to provide this breadth of support with only a few staff immediately available.

The shared options have not been developed in detail as a part of the NRAO computer plan described in this document. This section is an overview of the implications of sharing and our reasons for not pursuing this aspect further unless circumstances force us to do so.

4.2 - An NRAO Facility for Astronomical Computing

The option under which NRAO operates a facility which it must share with others is, purely from the NRAO point of view, the least bad of the four options because it would provide the greatest chance of success for NRAO's own objectives.

The disadvantages are obvious:

i) in either the near or far future the full capacity of the system might be needed for NRAO-related work;

ii) NRAO might have to support, to some degree, systems and applications software for very diverse purposes - theoretical number-crunching, NOAO, the Space Telescope Scientific Institute (STScI), and perhaps other federally funded university or national facilities;

iii) the administrative problems of such a unpredictably large group of users would be enormous.

To first order, the operational cost for a shared facility run by NRAO is likely to be double that for a purely NRAO facility. This cost is mainly for programmers and other support for software systems outside of NRAO's expertise. The extra cost could be even larger if one considers separate support subgroups for theoretical and non-radio astronomical computing.

The most serious problem would arise if the costs of the facility were to grow in such a way that the operation of the supercomputer would have to come, to some degree, at the expense of the NRAO budget. Clearly this problem would be much more manageable if the all users of the supercomputer were also NRAO users.

4.3 - Shared Facilities

NRAO's mandate goes beyond providing instruments and calibrated data, and includes providing the means for extracting the maximum amount of science from that data, i.e. providing the major source of computing. The computing needs are known to be large. If NRAO cannot obtain its own supercomputer, then it must consider sharing such a facility.

4.3.1 - Another NSF-Funded National Facility

Having another NSF-funded national facility for a landlord would be better than having other conceivable types of landlords, but sharing a computer facility under such an arrangement would have obvious problems. First, it might be difficult to get the landlord to be responsive to our needs. Second, the question of cost control, discussed in connection with the previous option, reappears under this option in obverse form. It might be necessary, for example, for NRAO to be charged a fixed percentage of costs under circumstances where NRAO has inadequate control over the origin of such costs.

Given these problems, it can be argued that no existing national facility should have primacy in a shared computer facility. If sharing a supercomputer with other NSF-funded entities is unavoidable, perhaps a separate institute with independent funding should be established. Under this arrangement, however, a facility essential to NRAO might encounter problems having nothing to do with NRAO.

4.3.2 - A University Facility

All of the disadvantages of using another national facility for large computing would be present under this option, with extra problems arising because the interests of any university are internally focused, whereas NRAO has a national constituency. The recent history of some university computing centers, where computing was made so expensive that their own faculty and staff could not afford it, is a clear warning that this option could have disastrous consequences for NRAO. The tendency for university people to think of national centers as rich would be difficult to avoid, so they are likely to be amongst the most difficult landlords of all.

We would require a stable operating environment, so sharing with a facility that is pursuing research by modifying the operating system regularly would have to be avoided.

4.3.3 - A Non NSF-Funded Facility

NRAO already has taken advantage of an offer from the Computing Division of the LANL to use a Cray-1 to gain some preliminary experience on the use of a supercomputer. It is easy to imagine turning that very informal and temporary arrangement into something more formal and long-term. In fact, LANL has indicated a willingness to consider a proposal for just such an arrangement and it is quite likely that something can be worked out for an interim period leading to the acquisition of NRAO's own machine.

For the long term however, this option would not be satisfactory. The first problem is that of security. Even though we are using an "open" machine the precautions that LANL must observe, while understandable, are nevertheless totally at variance with the NRAO mode of operation. Practices which are only a minor annoyance now would become intolerable when large numbers of astronomers without security clearances, visiting for short periods only, would need immediate access to the machine; these would be especially severe during the fairly frequent co-operative and exchange programs with foreign observatories. Fortunately, not all national facilities require security clearances.

In addition, providing a large fraction of NRAO's computer support in this way has the largest risks in the long run. There is no guarantee against shutdown of the facility by the dominant sources of funding. Their interests are even more different from those of NRAO than those of a university. For this reason the use of such a facility by NRAO, while convenient for now, must always be approached as an interim measure.

5. - A SPECIFIC PLAN

5.1 - Computer Hardware

Acquisition of a supercomputer by NRAO is the most direct and straightforward approach to its data processing problem since the required 60 to 80 Mflops sustained throughput cannot be achieved by any other hardware except through a great deal of replication. Furthermore, the dominance of array operations (such as FFTs) in the VLA task mix ensures that the special properties of a supercomputer can be exploited effectively.

The central aspect of NRAO's computer plan for 1985-1990 is the acquisition of a Class VI computer ("Supercomputer"). Even though supercomputer models are not numerous, there are still several choices that can be made. We have not pursued all possible options, nor have we been able to analyze completely those that we have investigated; we may well, therefore, not have found the optimum solution. However, we are confident that the configuration given here would satisfy our needs for performance and operating environment. It is therefore reasonable to base our estimates for capital and operating costs of a supercomputing facility on this one choice. The prices quoted are list prices in all cases; we have made no attempt to negotiate prices at this stage.

At full speed the Cray XMP will execute 220 Mflops but it is a rare application that can make use of that speed except in very small bursts. After initial setup a series of 2048 point complex FFTs can be performed on this machine at the rate of 1 millisecond each or roughly 130 Mflops. This is an upper limit to the rating of this machine vis a vis the VLA workload.

The word size in a Cray is 64 bits. The amount of main memory quoted above is therefore equivalent, in some ways, to 32 Mbytes. It should be noted however that the Cray instruction set does not include operations on units of less than 8 bytes. Despite the availability of utility routines to pack and unpack data, it seems clear that any smaller amounts of main memory would seriously compromise the machine's power to handle the large image problems for which it is intended.

A supercomputer can only obtain a reasonable fraction of its theoretical speed if the pipelines can be kept busy. Many operations involved in processing synthesis array data involve relatively few operations on a large amount of data, more data than can be kept in memory. Thus data and intermediate products must be kept on disk and read into the CPU at a sufficient speed to keep the pipelines busy. Since the speed of most supercomputers is very high, a comparable I/O achieved bandwidth is required.

Table 2

A SUPERCOMPUTER CONFIGURATION FOR NRAO

A Cray XMP model 1400 (Single processor, 4 Mwords memory)	$7,000,000
An SSD auxiliary storage unit with 32 Mwords of memory	$3,300,000
20 Gbytes of on-line auxiliary storage (disks).	$1,000,000
2 Input/Output Processors with 8 Mwords of buffer memory	$1,000,000
3 Peripheral interface machines (e.g. VAX or IBM 4341)	$ 600,000
3 Hyperchannel interface units	$ 200,000
6 High-density (6250 bpi) tape drives.	$ 300,000
64 Interactive terminals and ports directly connected plus 16 or more ports for remote access (dial-in)	$ 150,000
20 Personal computer (PC) workstations	$ 125,000
Interface to communication equipment	$ 75,000
Installation	$ 250,000
Contingency	$1,000,000

Total	$15,000,000

The SSD is needed to augment main memory because the type of work to be conducted on this machine is heavily dependant upon efficient I/O. Note that a single 2k x 2k image would occupy all of the main memory and that some current algorithms for image enhancement involve as many as ten different arrays of the same size as the image. To some extent the SSD acts as an extension to main memory but it is not addressable at the word level i.e. it is not an extension of main memory in the way IBM's LCS extended memory on the 360 series machines. The SSD is read and written as a super-fast I/O device.

The 8 Mwords (64 Mbytes) of buffer memory form a common pool of memory space accessible to both I/O processors. The main purpose of this memory is to minimize idle CPU time caused by I/O blockage but it can also be used for "memory resident" datasets i.e. as a type of fast auxiliary storage.

Almost all of the disks would be connected directly to the XMP rather than to the front-end machine. The Cray model DD49 disk units have a maximum capacity per drive of 1,200 Mbytes; 16 drives would be needed to amass the 20 Gbytes of storage needed by the VLA. The transfer rate of these drives is 10 Mbytes per second.

The role of the three small peripheral interface machines deserves special attention. It is anticipated that two of these would be installed in the vicinity of the supercomputer; one of these would be connected by high speed communication link to the third, which would be located at the other major NRAO site. These are not necessarily front-end machines in the conventional sense. They are intended to serve as an intelligent interface between the supercomputer and the graphics devices; they would also serve as terminal concentrators for the supercomputer

The tape drives and a disks would also be hung on the peripheral interface machine. Although Cray supplies a block multiplexor channel to control fast (200 ips), high density (6250 bpi), IBM compatible tape drives, the operating system that seems best for NRAO (CTSS) does not (at this time) support any tape units on the Cray itself.

The personal computer workstations are intended to supply the programming environment most conducive to high software productivity. It is anticipated that initial program development, documentation, screen editing for all purposes, electronic mail, etc., would be provided by the PCs rather than by the supercomputer. These workstations are, in effect, a distributed version of the front-end machine typically used with a supercomputer.

No provision has been made for optical (laser) storage devices. Although it is assumed that VLA data will be archived on laser disks within a year or two the cost of an appropriate reader should be absorbed by the contingency fund.

5.2 - Software Aspects

It will be useful to identify four different types of software. Even though such a division is bound to be somewhat arbitrary, it will highlight the areas from which the most important software requirements originate.

5.2.1 - Vendor-supplied operating system components

Under this heading are included such components as compilers, text editors, file management services, inter-machine communication facilities, and general utility programs. We have chosen CTSS to provide these functions. Under this operating system, the small computers used as peripheral interface units do not serve the traditional role of a front-end machine. They would not be expected to contain software accessible to the average user, to whom they are essentially transparent.

Nowadays the very best utility software is available on personal computers. By contrast the programming tools available on supercomputers are deficient. They are inadequate even by the standards most NRAO programmers currently enjoy on VAXes and PDP-11s. This plan envisages creation of an appropriate programming environment by using PCs as programmer workstations.

5.2.2 - Owner-supplied operating system enhancements

The cost of local system enhancements depends upon the cleanliness of the operating system organization, the number of hooks provided in the system, the amount and reliability of system documentation and the stability of the system.

The two main points of contact between the operating system and NRAO application software are likely to be i) database organization and manipulation and ii) service to special devices in real time.

The data typically obtained with NRAO instruments must be accessed efficiently for two different purposes: editing and gridding. These two processes impose somewhat conflicting requirements upon the organization of the files in which the data is stored. In one case the "time-baseline" order is appropriate, in the other the "u-v" order is preferred. Another type of conflict involves data identification. A cluster of 200 to 500 data values can be identified by a small packet of 8 to 10 values. If all data is stored in a single cluster a tremendous penalty is paid by those applications which require only a single value from each cluster. If data values are stored independently an overwhelmingly large overhead is required to associate an identifier packet with each data value.

In whichever way these, and other, conflicts are resolved, it is clear that NRAO applications require a rather special database organization. Use of that database requires a package of low-level routines that can be called from the higher-level application programs. These low-level routines must mesh well with the operating system if the entire NRAO software package is to have any hope of running efficiently.

Although each I/O processor on a Cray machine is a small minicomputer, programmable in its own right, Cray does NOT offer any software products for real time data acquisition nor does it support connections to non-standard I/O devices. The basic VLA process of transferring observations from the on-line machines to other computers for subsequent processing ("FILLing") could not be done in real time. It appears, however, that at least one Cray user (Grumman) has written software for direct on-line data acquisition assuring us that a real time FILLer is at least possible. The Grumman software might be available for use by NRAO.

5.2.3 - Applications Software

The third category of software consists of the major application programs which are developed and maintained by a cadre of professional programmers. These main line programs are the software that meets the basic objectives of the installation.

Application software currently in use by NRAO falls into five major categories:

i) On-line telescope control and data acquisition routines at all sites.

ii) The VLA calibration and editing package.

iii) The VLA Pipeline system for image formation.

iv) The AIPS package for image formation and enhancement.

v) Various software used for telescope data reduction at Tucson, Green Bank, and Charlottesville.

5.2.3.1 - On-line Systems

The plan to acquire a supercomputer does not immediately affect the first category of application programs. If the supercomputer is purchased and located close to the machines used to control the VLA telescope, then it might be possible to link the two devices for a direct transfer of data from one to the other (on-line FILLing). The present plan does NOT include that connection as a critical element - only as a possible future convenience.

Plans also exist for moving some software from the DEC-10 at the VLA into the on-line system. This shift will reduce the need for real time FILLing.

5.2.3.2 - VLA Calibration and Editing

The VLA calibration and editing package will require more conversion work than any other NRAO software. It is written in a non-transportable language (SAIL); the database structure cannot easily be expanded to fulfill our existing requirements; it has been maintained for so long that the structure of many major components is buried; and finally, the original design exploited, to the point of abuse, a high-level language feature (macros) that is not found in any other modern programming language.

It is clear that the VLA calibration and editing package will have to be rewritten. Since the equivalent programs for the VLBA must be created at this time, it is opportune to include both in a single package. Whether or not to cast this integrated package in the framework of AIPS or to preserve the "Standard Command" mode of input currently in use on the DEC-10 is still being discussed. In either case this package should be ready to go on the supercomputer by the time the machine becomes available. Further discussion of this point is presented below under the heading of Transitional Activities. It should be noted however, that the upgrade of the VLA calibration and editing package is needed, and will proceed, regardless of the outcome of the plan to acquire a supercomputer.

5.2.3.3 - The Pipeline

The Pipeline system is overly complex, too fragile and too difficult to maintain. It will simply be abandoned as soon as the supercomputer is available to carry the load.

5.2.3.4 - AIPS

The feasibility of this plan for NRAO to use a supercomputer is enhanced by the fact that a significant effort has already been expended during the past few years to make truly portable a major portion of its application software (the AIPS system).

AIPS has been ported to several different computer hardware architectures other than the VAXes on which it is used most widely. These other machines include ModComp, IBM-370, CDC 60-bit, and Motorola 68000. It has also run under more than one operating system: VMS and Unix on a VAX; OS/360 and Unix on IBM. The portability of AIPS is due to the use of a portable dialect of Fortran, highly modular coding, isolation of machine- and Operating System-dependent code, and parameterization of system and device characteristics

Although the AP microcode is a critical element in the success of AIPS as a total system, this microcode has been carefully confined to a limited number of modules, which makes it possible to replace these modules when implementing AIPS on a new kinds of vector processing hardware i.e. on a supercomputer.

The AIPS system will be used in the supercomputer benchmarking effort; it will probably be made available to other supercomputer users (e.g. NASA) whether or not NRAO ever acquires a supercomputer of its own. If so, the AIPS system would be available for use as soon as NRAO's supercomputer became operational. NRAO will have to support software systems for these smaller systems as well as for the big machine for two main reasons:

i) Network access to the central supercomputer is unlikely to be good enough everyone. Users will still want to take their data home and continue analysis there. In the end, there is really no substitute for having your own computer!

ii) It seems very likely that NRAO's own superminicomputers will survive for some years, used more for image analysis than for pure mapping.

NRAO's users will want to have the same command language and application tasks available at home as at NRAO sites. Development of the supercomputer software should be done with downward compatibility, in the same sense that the existing AIPS code has upward compatibility to the supercomputers.

5.2.3.5 - Tucson, Green Bank and Charlottesville Software

It is clear that the impact of the planned supercomputer will be greatest on the VLA and VLBA sites and on AIPS users. This plan does not address the problem of single dish software or general purpose utility software; some will undoubtedly be rewritten for the large machine. Some of the large programs will not be converted (e.g. the structural engineering packages used for antenna design, which are IBM specific), and computer time for these will have to be rented on an outside machine if equivalent programs are not available. The hardware plan does include sufficient communication equipment to enable NRAO users at other sites to process their data via the supercomputer.

5.2.4 - Developmental and Casual Software

Programs developed by the general computer user form the fourth, and final, category of software. These one-shot programs are generally much smaller than the main line programs described above, but even trivial programs can contribute significantly to scientific productivity by meeting immediate individual needs.

It is fairly common for scientists to find that they are spending a large fraction of their time writing computer programs. In order to avoid the expenditure of a disproportionate amount of professional scientific time on amateur programming activities, the legitimate need for casual programming must be satisfied by making available the most widely known languages, a well documented library, and a "user-friendly" operating system.

A different view of casual programming is obtained by considering two basic facts: i) a very large fraction of what starts out to be a private programming endeavor ends up being adopted for wider use and ii) almost all code produced by programmers is ultimately maintained by other programmers. From these facts it follows that NRAO certainly has a right to be concerned about the choice of languages in ALL programming projects and that management concerns about code quality as well as language portability are not really invasions of the intellectual rights of either scientists or professional programmers.

5.2.5 - The Question of Programming Languages

The NRAO plan for programming language for the supercomputer is to use the most up-to-date Fortran wherever possible, consistent with maximum transportability of those software elements for which transportability is important.

When, and if, Ada is widely implemented and becomes available for use on the NRAO supercomputer, a re-evaluation of the language question will be in order.

In the on-line systems, assembly language will be used only for those components of where speed is critical or where the Fortran language constructs are inadequate. For those on-line systems that are running under Unix, the C language will be preferred over assembly language. Other than that, the C language will not be used internally but it may be used in software systems that will be exported (e.g. a Unix version of AIPS).

5.2.6 - Other Comments on Software

The productivity of the professional programming staff will depend heavily on what is now called "the programming environment." It includes purely software tools such as editors and interactive debuggers, management tools such as a program librarian and documentation aids, and hardware tools such as personal computers and modems for convenient (home) access to central facilities.

6. - OTHER CONSIDERATIONS

6.1 - Operating a Supercomputer

NRAO is proposing a large-scale computing facility; it will necessarily become a large-scale operation, consuming a significant fraction of the available NRAO physical, monetary and manpower resources. This section outlines some plans for operating such a computing facility in order to maximize the science performed, provide the desired service to users and remain flexible as demands and budgets change with time.

There are a number of assumptions built in to this discussion:

i) NRAO will own and operate its own supercomputer facility: it will not be shared or serve as a national center for astronomy computing;

ii) NRAO will cease to operate some of the present synthesis mapping facilities. The VLA DEC-10 and PDP-11's and the Charlottesville IBM and ModComp will be replaced by the new facility. The VAX systems will be retained to serve as front-ends to the new facility, as stand alone image processors, or both.

iii) Some NRAO computer manpower will be relocated to the site of the new facility.

Under these assumptions, operation of a supercomputer facility would be much more straightforward than either a shared facility or the current NRAO systems. There will be one system to support and maintain and it will be under the control of NRAO. If any of these is not true, more manpower will be needed than discussed below, especially during initial installation.

6.1.1 - Physical Plant

A supercomputer requires a great deal of power and air conditioning as well as structurally sound floors and damage protection systems. The computer room must have, at a minimum:

Adequate floor space for the CPU, peripherals and front-ends
Motor-generator set(s) and power plant room(s)
High capacity air conditioners with filtration and humidity control
Cooling water supply with piping
False floors for cabling
Hardware maintenance office and repair shops
Fire protection system (Halon)

Estimates of requirements can be made from discussion with the system vendors but are likely to exceed present capacities, particularly for power and refrigerant - NRAO does not now have a computer room with the required facilities. Detailed site planning must be done in consultation with vendor as early as possible. The requirements for the new computing facility, if it is to be in Socorro, will affect the design of the VLBA operations center. To house the computers, new building space and environmental equipment will be needed as follows:

Table 3

Capital Expenses

(1984 $)

	Area (sq. ft.)	Unit cost ($)	Total cost ($k)
Building space for:			
CPUs and Peripherals	2,500	75	187.5
Maintenance Shop	500	75	37.5
Additional Office Space	2,000	60	120
Ancilliary Equipment (below)	500	60	30
Cooling (100 Tons)	-	-	90
Spares	-	-	15
Power Conditioning (125 kVA)	-	-	65
Spares	-	-	10
Fire Protection (Halon)	-	-	25
	5,500	105	580

6.1.2 - Hardware Maintenance

Since a supercomputer is a very special machine, hardware maintenance is ordinarily left to the machine vendor by contract. Supercomputers are expensive maintain; a service contract for a typical Cray-1 system, for example, costs in the neighborhood of $70,000. per month, about 5 times the current DEC maintenance contract at the VLA alone. However, the service is a complete one, covering ALL peripherals and ancilliary hardware 24 hours a day, 7 days a week with at least one resident service engineer.

Necessarily some equipment will be from a third party or home made, but this should be kept to an absolute minimum. In maintaining computer systems, NRAO has come to the (hardly surprising) conclusion that add-on hardware from third party vendors is not as reliable as that provided by the mainframe manufacturer, even though the initial cost is lower. Hardware maintenance on multi-vendor systems has been shown from experience to be expensive in terms of manpower and system downtime and will become even more important when the new facility is the ONLY way to process the larger problems; there will be no equivalent system to take over when the main machine is down for repair.

6.1.3 - Annual Budget

Operation of a supercomputer will cost much more than even a large mainframe and more than the combined costs for running NRAO's present computers. However, increasing compute power by an order of magnitude costs much less than an order of magnitude more to operate. Table 4 shows a rough estimate of the increase required in NRAO's annual computing budget in order to replace the DEC-10 and PDP-11 systems at the VLA and the IBM and ModComp in Charlottesville with a supercomputer facility.

Table 4

Annual Operating Budget

(1984 $k)

Category	Current	Increment	New Total
Capital Investment	500	900	1,400
Contract Maintenance	600	300	900
Communications	0	200	200
Materials and Supplies	200	50	250
Salaries and Benefits	1,500	625	2,125
Travel	75	25	100
	2,875	2,100	4,975

The increase in the investment budget is based on the assumption that it should continue to be about 8% of capital investment in order to minimize obsolescence. This means $1.2 million for the proposed supercomputer facility alone; additional capital money will always be needed for the other computer systems at NRAO. Although this seems an enormous sum, the equipment that would come under this heading is very expensive (e.g. an on-line common file system with its attendant computer). These figures are obviously subject to discussion based upon the relative importance of other capital items (e.g. buildings, electronics, etc.) required by the observatory.

The M & S budget assumes that the availability of larger numbers of larger maps will cause a sufficiently large growth in long-term storage requirements that we will loose ground in the battle to reduce storage costs, i.e. that increases in storage efficiency will NOT keep pace with the increase in demand.

In addition an estimated 250 kW power is necessary to operate the supercomputer center, which at present rates would cost about $150k per year. We assume that about 30% would be saved by ceasing to operate some existing equipment. Thus, NRAO should budget for an increase in the power bill of $100k.

6.2 - Manpower Needed

It is hoped that NRAO can operate a supercomputer center with fewer personnel than is customary at other such installations. The typical supercomputer facility has a large base of relatively inexperienced programmers all attempting to use the machine, debug code, and optimize performance. The situation for NRAO would be quite different; a small group of expert programmers would develop and maintain most of the software (as is done now, for example, with the DEC-10 calibration package or the AIPS reduction systems). Users would be just that - users of existing software packages. Hence, the need for a large number of user assistants would not exist.

Table 5

Manpower

Area	CV now	VLA now	TOTAL now	Needed	Incr
Management	1	2	3	4	1
Secretarial	0	0	0	2	2
Systems Analysts	2	1	3	6	3
Prog. Analysts	6	6	12	12	0
New Project Progs.	0	0	0	3	3
Math. Analysts	1	1	2	4	2
Hardware	1	2	3	4	1
Comm. Specialist	0	0	0	1	1
Operation Support	2	2	4	4	0
Operators	5	1	6	8	2
	18	15	33	48	15

Notes:

i) 13 people involved in other NRAO computer activities (i.e. on-line systems, single dish software, and VLBI) have not been included in the above list.

ii) The Charlottesville operators currently serve the VLBI equipment as well as the computers. It is assumed that the former activity will cease when the VLBA becomes operational.

iii) The increase in the number of mathematical analysts is not strictly necessary for the purchase of a supercomputer, but the computer performance would very likely be enormously enhanced by such analytical development (see section 6.5.6).

iv) New instrumentation planned, such as the 75MHz array, will certainly produce a demand for more programmers, but not necessarily in the first years' operation of the supercomputer.

6.3 - Communications

The requirements for communications, between a large NRAO computer installation and other NRAO sites, are currently less certain that the requirements for computer capacity. Considerable effort has gone into an objective evaluation of the latter, based on scientific potential. The communication requirements, however, are much less developed. Whereas the computer capacity sets a fundamental limit, there are probably a number of different but acceptable working communication solutions.

We would like to make NRAO's major computer facilities available to all NRAO users and staff. Ideally one would like full capabilities everywhere; that is, the user's capabilities would be independent of his location relative to the computers.

6.3.1 - Geography and Required Capabilities

6.3.1.1 - NRAO-Based Data Processing

The location of the computer center (assuming only one system) has not yet been addressed. Assuming an NRAO facility, the most likely candidates are: the VLBA operations center in Socorro, the NRAO headquarters in Charlottesville, and the VLA site. The VLA site has the advantage of being the source of the bulk of the data and the disadvantage of being somewhat remote. A decision as to the eventual location involves complex issues outside the scope of this document. However one possible scenario, with the computer center in Socorro, will be used here as an illustration. Fortunately, the nature of the communications problem is such that the details of this scenario will change very little with other alternatives.

High speed access to the facility is needed from the Charlottesville headquarters and the VLA site. In this context high speed means at least a T1 carrier (1.5 megabaud) capability so that images of 1k x 1k pixels, with one byte per pixel, can be transmitted for display purposes in less than 15 seconds. It is unclear whether the norm would be to transmit a map of one or two bytes per pixel for display or two or four bytes per pixel for further local processing. In some cases transmission of visibility data will be required, but the display oriented transmission will always set the most stringent time scale. Using the above communications rate, a 4 million point (4 bytes per pixel) map could be transmitted in a time of 4 minutes. A longer transmission time might be acceptable if the transmission could be overlapped with local processing.

The Green Bank and Tucson facilities should also have access to the computer facility. Since these sites have, and will continue to have, their own computers, most of their initial requirements can be satisfied by a 9600 baud link, similar to the existing NRAO inter-site digital communications system. This can be upgraded as necessary.

6.3.1.2 - Raw Data Transmission

The above discussion addresses only the capacity required for interactive data processing. It must also be possible to transmit data directly from the VLA on-line computer system to the major computer facility. The worst case continuous load currently envisioned is 560 kbaud (i.e. 351 baselines x 512 spectral channels x 32 bits per complex number every 10 seconds). At present, only 35 kbaud is possible (40 x 256 x 32 / 10). A communications link with capacity at least that of a T1 carrier should be planned if this data rate were to be needed. Fortunately, in almost all cases the data rate can be reduced either by increasing the integration time, by transmitting only a small percentage (calibration data - 15%) in real time, or by incurring a several hour delay with magnetic tape transport.

The current plans call for the operation of the VLA to be moved to Socorro. If this is done, more real time diagnostic data will be needed from the site. At present, such data is communicated at about 35 kbaud.

It seems likely that a more modest transmission facility of about 500 kbaud could therefore be used to fulfill the real time data acquisition requirements.

6.3.1.3 - University Use

It is difficult to specify the required capacity or capability for remote user use at a typical university. Ideally, one would like a high bandwidth capacity to each of some 20 to 30 universities; this is hardly practical or affordable. A 1200 baud link, for program submission, exists at present (on a dial-in basis). This, however, is not considered adequate by many university users. Good communications to the university community is necessary for two reasons.

i) During data processing, the astronomer typically needs to see the results of one processing step before going on to the next.

ii) The diversity brought about by the mix of university users will probably be necessary to load-balance the machine. The number of resident users will frequently be too small for good load-balancing. A system with most users at the end of a 1200 baud link might not be sufficient to elicit a adequate level of outside use.

It is recognized that users will usually need a wide bandwidth only for receiving data from the central computer. After studying the images, the user could use a 1200 baud link to send command sequences back to the computer. It would be possible for the main computer to transmit images via satellite to a receive-only station in a university. It was hoped that such a receive only-facility would cost $30k, with an operational cost of under $1k per month, which might be acceptable. Early estimates however are unfortunately much higher. If it is assumed that university facilities will contain superminicomputers capable of some level of image processing, then the requirement for the transmission bandwidth of the link to the main computer could probably be reduced to a more realistic 56 kbaud.

The NSF, under its Large Scale Computer Initiative, is known to be planning a scientific communications network. Initially the proposal is to inter-connect some 20 universities and large computer centers with bandwidths in the 50 kilobaud to several megabaud range. It is possible that NRAO's large computer facility would be a node on this system. The system being discussed by NSF is planned to expand to include access by a significant part of the university community. Exactly how much of NRAO's problem would be solved by this system will not be clear until more information is known about the NSF project. However, we could plan our system with links to NRAO sites as discussed above, and assume that the proposed NSF-funded system will eventually be the NRAO-university link.

6.3.2 - Technet Project

The state of New Mexico has allocated $12 million to build a high bandwidth network connecting a number of large computer facilities in the state. This network is proposed to be run by a nonprofit corporation using both fiber optic and microwave technology. During the initial phase, the link will connect only Los Alamos with Albuquerque, but the later phases of the project will connect these with the White Sands Missile range. At that time, NRAO could be connected on to the link to provide the communication capacities between the VLA site and Socorro and the other supercomputers in New Mexico, giving as well a wide bandwidth link to ARPANET.

6.3.3 - NRAO's Existing System

NRAO currently operates a digital network among all sites. The system multiplexes leased lines between the sites using multiple protocols. DDCMP (for DECnet), SDLC, BSC and asynchronous transmission are used as separate logical links on single physical links. DECnet is used for inter-site electronic mail, and other data exchange, using these links. This digital system has been very successful and has had considerable influence on NRAO technical operations. It is therefore obvious that this inter-site communication system would be one of the minor uses of the wideband communications links that we are discussing.

6.3.4 - VLBA Communications System

The VLBA project plans to develop a communications system to connect the Socorro operations center with each of the 10 antennas in the continental USA, Hawaii and Puerto Rico. The system will be narrow band and will, at least at the start, make use of leased phone lines in a manner similar to that used by the above described NRAO inter-site system. Serious consideration is being given to the eventual use of satellite links.

6.3.5 - Budget

For budget purposes, it is assumed the computer is located in Socorro. 1.5 Mbaud satellite links are assumed to connect the VLA site and Charlottesville with Socorro. Universities are assumed to be connected to the NRAO computer via a satellite wideband receive-only links, the reverse path being over phone lines.

A connection to the NSF-supported Science Net is assumed. However, its effect is assumed to increase the number of accessible universities rather than to decrease the NRAO's direct cost.

Table 6

Communication Costs

Description	Capital ($k)	Operations ($k/yr)
Satellite Ground Equipment		
Charlottesville	250	20
Socorro	300	25
VLA Site	250	20
Satellite Channels (2 transponders only)		145
	---	---
	800	210
For university receiving sites:		
Receive-only system	50	7
Dial-in facilities	25	<3

6.4 - Image Display Facilities

The main objective of the NRAO computing system is to provide the mapping throughput necessary to operate the VLA at full efficiency, including the full spectral line capacity designed into the hardware. The most important factor governing its throughput is obviously the computer capacity as measured in I/O bandwidth and in compute cycles. While the major concern in the formulation of the computer plan has been to provide this capacity this alone is not enough to make it successful. If a supercomputer is to have its full impact on the amount and the quality of science done with the VLA the astronomer also needs adequate display tools to enable him to digest and interpret his data as easily and efficiently as possible. Since decisions on subsequent processing steps are often based on an assessment of the data and the objectives, an interactive system is best.

Two important components of data display are graphics (line-drawing) and image (grey-scale) display; both may be in monochrome or in color. There is an obvious difference in application: identical data may be displayed on the former as a graph or contour map, on the latter as a "photograph". Such displays are complementary.

6.4.1 - The image plane

Foremost in importance is the ability to display the final images. The display of a single image of 512 x 512 pixels, which is the typical size used in the existing systems, presents no problems. Image display devices with storage for 2048 x 2048 pixels using monitors with 1024 lines are available.

The main objectives of image display are to give the astronomer an instant overview of the contents of his data, as well as to enable him to lift out details. This requires high flexibility and interactivity in the process of translating the bits to an image. The basic tools here are color schemes, transfer functions, zoom and pan, selection of image segments, and comparison of images.

In addition, there is the problem of the third dimension. The VLA is capable of producing multi-channel images or "cubes" up to 2k x 2k x 512 pixels but there is also a requirement for smaller cubes at various stages of processing. The Image Storage Unit currently under development at NRAO has a capacity of 512 x 512 x 512 and will provide some of these capabilities, but for the longer term even more sophisticated hardware for the selection and display of interesting data is required. Three-dimensional display systems will provide some relief here. Three-dimensionsal graphical display systems are already available but in the image display area we will have to turn to stereoscopic and animated displays and perhaps develop holographic display techniques.

6.4.2 - Visibility plane

This is the Fourier transform of the image plane, and the one in which the measurements are made. The visibility displays are used to monitor the calibration quality and to recognize and edit corrupt data. Again, an interactive environment is highly desirable. The data comprises complex data (i.e. real and imaginary), and at least two coordinate systems are necessary; the first is the natural ordering in which the data is collected (known as baseline-time order) and the other the spatial frequency ordering (known as u,v-order). Some instrumental error patterns are only contiguous, and hence easily recognized, in the baseline-time display while other systematic sky based structures will only appear in the u,v display. Both are important, but it requires a complete sort to convert between the two orderings.

6.4.3 - Image Analysis

Modern image display devices are able to perform simple image analysis functions such as image statistics, two-dimensional convolutions, etc. Although these operations are performed with limited accuracy they are done at interactive speeds without requiring (much) computational and I/O support from the host. We should take full advantage of these capabilities because they provide to the astronomer a powerful and fast experimentation and preview tool.

6.4.4 - Hard Copy

Hard copy of the images is also necessary for both working and final copies.

i) Black and white or color copies of the displayed image to be obtained from the monitor video signal.

ii) High quality images to be written on an image recorder. This has the advantage of high capacity and good quality. For example 64 512 x 512 images can be recorded in pseudo-color, using a Dicomed film recorder, on one film sheet with no loss in resolution.

6.4.5 - Vector Graphics

Two-dimensional line graphics, e.g. contour maps and cross section displays, are inadequate to convey much information about a large image but they are essential to provide quantitative information about smaller regions in the image. The line and point graphics applications are not much of a technical problem. GKS (Graphical Kernel System) offers the possibility that we can provide device independence in our graphics software. There are also good software packages commercially available that are compatible with GKS.

6.4.6 - Display work station budget

Proposed are three classes of work stations: large, medium, and off-line. Large work stations comprise a three-dimensional graphics system (say, Evans & Sutherland), an image display system (say, IIS), peripherals (hardcopy, video cassette recorder with editor, etc), and in-house hardware (image storage device, control panel). Medium work stations are identical, except that the three-dimensional graphics device is replaced by a simple raster scan graphics device. Off-line work stations consist of hardware intended for viewing images and/or animations already produced. The estimated budget is given in Table 7.

Table 7

Display Budget

(1984 $k)

Number	Item	Component	Cost	Existing Hw.	Net Cost
2	Large WS	3-D Graphics	80		
		Image Display	55		
		Peripherals	35		
		In-House Hardware	30		
			200		400
3	Medium WS	Graphics	20		
		Image Display	55*		
		Peripherals	35*		
		In-House Hardware	30*		
			140	-120	60
4	Off-Line WS		20		80
2	Image Recorders		100*	-100	0
	Total Cost				540

*Hardware already exists.

6.5 - The Computing Milieu

The following is a discussion of selected management issues that have an important, albeit indirect, influence on the effectiveness with which computers are used at NRAO. These issues generally relate to the working environment in the NRAO computer department but they have additional importance outside of that area.

6.5.1 - Management

The charter of NRAO is to support basic research in radio astronomy. NRAO's size is small by government-funded laboratory standards, and appears even smaller because of its geographical distribution. The basic research orientation and small size lead to a rather uncommon style.

Authority tends more to follow interpersonal relationships than assigned structures. Development projects frequently involve mixtures of professionals from technical divisions and scientists from the basic research division. Frequently the individual intellectually best suited to develop a project is not well suited to take management responsibility for the project. As such, responsibilities do not always end up as clear cut as they might be with some decisions made by scientific staff and others by technical staff. Although strange, this approach has been extremely successful in much of the Observatory's efforts. The development of the large system under consideration will probably force a much more formal management approach at least on that portion of the organization.

6.5.2 - Accounting vs Open Access

Computer accounting systems are used for four purposes: to distribute fairly the use of the facility, to distribute costs among various budgets, to increase efficiency by tuning the system to a given mix of applications, and to discharge reporting responsibilities.

NRAO has traditionally avoided the use of accounting systems as a means of rationing computer use. In the past it has successfully relied on peer pressure to avoid major abuses. The most visible instances of this practice have controlled utilization of disk space on the DEC-10 and AIPS machines, both chronically scarce resources. A related technique is the use of a sign-up sheet for AIPS processing time on the VAX computers, which effectively allocates the use of the AP and the image display device on these systems. This practice does not seem to have caused undue strain in the organization.

In general, the computing environment at NRAO is an exceedingly benign one. A login procedure is used to prevent accidental interference between users and also to protect against totally unauthorized access via telecommunication. Further security methods have never been implemented because it has never been a problem.

Since NRAO serves a single sponsor (NSF) with the single broad mandate to conduct research in radio astronomy and a small user community whose interest is the same, there has not been much point to an internal accounting system so elaborate that each scientist or staff member was allocated his or her own computer budget.

None of this means that no one is paying attention to the purposes for which the computer is used. Statistics on the amount of computer resource devoted to development as opposed to production and to NRAO staff as opposed to visiting scientists, are collected and reported.

The current mode of computer operation by NRAO has important advantages that should be carefully preserved. One clear advantage is reduction of overhead costs. This, however, is insignificant compared to the productivity gain that comes from uninhibited access to one of the most important research tools. The very worst result that could ensue from the installation of a supercomputer at NRAO would be the stifling of creativity by bureaucratic fences (e.g. refereeing of computer use). Nevertheless, it is recognized that some additional control shall be necessary.

6.5.3 - Open shop vs Operator Assistance

No one would propose turning over to the visiting observers control of the VLA telescope. It is quite obvious that human safety as well as efficiency of use requires that operation of the instrument be entrusted to highly trained specialists. The situation with regard to computers at NRAO is not quite so clear cut because human life is not (generally) at stake but the question of effective use of a costly, complex and delicate piece of equipment is the same.

In the past one of the NRAO computer facilities has been run with the assistance of operators (the Charlottesville IBM machine) while all others have been run as an "open shop." Interestingly enough both practices have had to be modified. Since the operators of the IBM computer are frequently busy with other assigned tasks, users of that machine are allowed to perform some operator functions themselves, e.g. mount their own tapes. At the VLA, on the other hand, the addition of some operational has been necessary.

To some extent NRAO telescope operators have been acting as computer operators. Of course they have full operational responsibility for the on-line machines. But in addition, at the VLA, the telescope operators reboot all machines after power failures or other malfunctions; upon request they will occasionally mount a tape for a remote user; they will replenish the paper supply when it runs out at night, change ribbons etc.; they do some "hand holding" for inexperienced users; they know who to call when serious problems arise; and, in general, they keep a watchful eye on the computers as well as on other equipment.

In general it appears that, regardless of the type of equipment in use, computer operators are needed under two circumstances: when the hardware is being used from remote sites and when many users are working simultaneously on a shared machine. Since both of these conditions will apply to a supercomputer facility it is fairly clear that a full-time staff of computer operators will be required.

6.5.4 - Professional Programmers

Most software produced at the NRAO has been developed by astronomers, in many cases astronomers turned programmers. This is a result of many factors and probably follows a pattern similar in many basic research organizations. The organization benefits in many ways. One frequently finds a more intelligent individual, who requires less salary, and who better understands the problem. It should be emphasized, however, that there are dangers if this approach is carried too far. If a proper mix of professionally trained programmers is not maintained, the software will lack professionalism.

6.5.5 - Multiple Locations

The physical separation of the various NRAO sites and the corresponding fragmentation of the computer staff greatly increases our problems. In addition to making management more difficult, the separation also complicates the technical activities. The problem of developing a software package that will be heavily used at two different places is a difficult one. Joint development over a phone line is not easy. Development of the program by one group usually leads to software considered ill suited by the other. Past analysis of problems in this area usually have drawn the conclusion that each group did not understand the special problems and desires of the other. There are other examples also where the distance effect is detrimental. The effects of a move from the NRAO's current somewhat distributed computer hardware to a single large facility are not clear. Supercomputers are normally used by programmers at distributed locations. In most cases however, the programmers are working on totally or nearly independent projects rather than on an integrated package.

6.5.6 - Future Projects Group

NRAO has an engineering development laboratory with the responsibility to research and develop new instrumentation for NRAO. This has been extremely valuable in the production of improved receivers and the opening of new frequency windows. Aperture synthesis depends on computer technology as much as on electronics technology. Improvements in computing capability are likely to come from improved algorithms as well as from improved hardware.

To illustrate the need for continued image processing research we will briefly review part of the history of image processing in radio synthetic imaging. The CLEAN algorithm was originally developed for the iterative removal of point sources from a synthesized image. Since the number of point sources removed was typically small, the computer time required for a substantial improvement in the image was relatively modest. However, after some use in this restricted fashion, it became clear that CLEAN could also be used to deconvolve images of quite complicated sources, which were not obviously the sum of point sources. Of course, the required computer time increased by a large factor because the number of point sources removed had increased by up to two orders of magnitude. At the end of the last decade, Barry Clark at the VLA invented a variant of the CLEAN algorithm which exploits the AP-120B to achieve a performance increase of up to an order of magnitude. More recently, Bill Cotton in Charlottesville developed another variant of the CLEAN algorithm for the AIPS system which is based on direct subtraction of point sources from the observed data, and which can gain about another order of magnitude in speed in some circumstances. Another deconvolution algorithm, the Maximum Entropy method, has also been coded in AIPS, and can provide superior results on the larger sources, with substantially less computing time. From this history, we can see that the developing hardware, software systems, and algorithms are all interlocked and interdependent. New hardware often permits work on new algorithms, and new algorithms often require new hardware. The net result of the interplay between these three elements (hardware, algorithms, and portable software system) is, in the end, better science.

While both the hardware and software systems have been the focus of considerable effort at NRAO, the development of new image processing algorithms like CLEAN has been mainly due to individuals scattered over the Observatory. It has been suggested that NRAO should establish a group with a mandate to coordinate the investigation of new computer hardware developments, software systems, and image processing algorithms. To a large extent, the people would probably be those on the current NRAO staff who have worked in this areas. The installation of a supercomputer at NRAO would naturally lead to all of the algorithm research people working on the same machine. The shared access to common data and common algorithms through the wideband link would naturally stimulate co-operative research projects, even among individuals at different NRAO sites (we already see this trend with our existing low bandwidth network). It is obvious that research interactions would be optimized if all of the relevant people were located at one NRAO site.

If such a group were formed, it should specify how the exploration of new image processing algorithms would be related to the development of actual working systems; it should also specify policies to be followed in hardware development projects. It should avoid hardware development which duplicates existing technology, and instead concentrate on the evaluation of commercial systems and on the development of innovative combinations of available systems.

The possibility of such a center is discussed further in VLA Scientific Memorandum No. 144 ("Image Reconstruction at NRAO").

7. - TRANSITIONAL ACTIVITIES

7.1 - Hardware

There is a staged project, already underway, to upgrade the VLA on-line computer system by replacing some of the ModComp minicomputers by more modern machines from the same vendor. The first purchase has already occurred; the second is planned for two years hence. Since there appears to be no conflict between that activity and any possible outcome of the longer term plan presented in this document the upgrade of the on-line system should continue while attempts are being made to bring this plan to fruition.

The most pressing need for off-line computing is a replacement for the ModComp (in Charlottesville) that was used as a development machine for the AIPS project. Some funds will probably have to be expended for mini- (or micro-) computer hardware for this purpose before the fate of this proposal is known. As in the previous case, there appears to be very little conflict between the objective of the proposed expenditure and this plan. Regardless of the outcome of the plan proposed herein, NRAO will remain committed to provide software support for AIPS systems running at small university installations and that will be the purpose of this machine.

Facilities for reduction of actual observations is painfully slow at Charlottesville. This situation may be remedied by the installation of a super AP (such as a Star ST-100 or a Numerix Mars-432) on the IBM 4341 already in use at that site.

The relatively modest needs at other NRAO sites have been satisfied quite recently by the purchase of two VAX computers for Tucson and a Masscomp system for Green Bank.

7.2 - Software

Programming during the transition period leading to the acquisition of a supercomputer can be roughly divided into the following four areas:

i) A major upgrade of the VLA on-line control and monitor system. Creation of equivalent software for the VLBA.

ii) Continuing development of site-specific software at Tucson and at Green Bank.

iii) A total re-write of the VLA editing and calibration software. The new version will incorporate features required for VLBA data reduction.

iv) Low level maintenance of both the Pipeline and the VAX-based image formation and image processing systems.

v) Benchmarking. Intense activity by few people over a short interval to mostly to rewrite the few machine-dependent routines necessary for AIPS to run on a Cray supercomputer under CTSS.

7.2.1 - VLA on-line control and monitor

The on-line hardware system comprises ModComp computers for which the only programming language available for time critical tasks is a macro assembler. About 80% of the existing system is written in assembler, the rest is in Fortran 66. With the advent of the new ModComp computers, a much greater fraction of the code will be written in Fortran. On the ModComps no other realistic choice exists for an alternative high level language.

The standard which has been adopted is Fortran 77 with only a few extensions which have been deemed essential for good project management (e.g. the INCLUDE directive to ensure commonality of COMMON areas, array sizes and variable names).

7.2.2 - The VLA and VLBA editing and calibration package

The software addressed here is written in SAIL for the DEC-10 machine running under the TOPS10 operating system. The SAIL package has a limited future because:

i) It is written for the 36-bit hardware, which has a limited future because DEC is not providing an upgrade path for DEC-10s or DEC-20s and no other major vendor is considering such a memory architecture. The programs are strongly tailored to take advantage of the 36-bit word size.

ii) The SAIL code, as written, is difficult to migrate to any other machine (including a DEC-20). The main problems are abusive nesting of the high-level-language macros (a very useful but unique feature of SAIL) and the dependency on the TOPS-10 architecture. SAIL is designed to make many system calls look as close as possible to the TOPS-10 assembler system calls.

iii) SAIL as a language is only supported by one small company.

The existing SAIL package functions well, but needs extensive modification for proper spectral line support, removal of the continuum kludge, and increased graphics support. A complete set of library routines now exists so that the package could be totally rewritten in Fortran, but those library routines assume that the current inadequate data structures will be preserved.

In addition, the calibration software for the VLBA project must be written. Here, the basic functions are now available under AIPS, but currently not with the desired flexibility. There is therefore a proposal to combine the VLA and VLBA calibration and editing software in a major new package in Fortran probably using AIPS.

Some of the existing calibration and editing procedures for the VLA will be performed automatically by the new on-line system.

7.2.3 - Maintenance of AIPS and the Pipeline

The major image formation systems are the Pipeline and AIPS. The software written for the Pipeline is targeted very strongly toward the efficient use of the existing hardware and is consequently not transportable to any future system. The AIPS package will remain the basis for future image processing systems. There will be a major effort to port this to a supercomputer in the same way as it has been ported to other operating systems than its native VMS (e.g. Unix). The standard language may be upgraded to Fortran 77 and another command language interface (e.g. IRAF) may be considered.

In general, the programmers who have been working on these two projects will be diverted to benchmarking and other supercomputer related activities. The amount of work that will be done directly on these two projects will be relatively small during the transition period. It should be noted however that i) conversion of AIPS to run on a supercomputer will yield indirect benefits to the overall AIPS package and ii) the rewrite of VLA calibration and editing (including VLBA aspects) might also be done under the AIPS umbrella.

7.3 - Benchmarking

Under a temporary and informal arrangement with the LANL, an NRAO summer student working with one of the regular NRAO staff has been using a Cray-1 to do the deconvolution of a high resolution version (4k x 4k) of the image described in the introduction (Figure 5). This project has the dual objective of giving someone in NRAO firsthand experience with a supercomputer, and producing a scientifically valuable result that could not be obtained except with great difficulty on the VAX machines. From this project we have already gained some insight into the problems and the costs of dealing with large datasets on a supercomputer.

7.3.1 - Running AIPS on a Supercomputer

It is our intention to convert the few, clearly identified, machine dependent routines in the AIPS package to run on a Cray computer under the CTSS operating system. This project should require 4 to 6 man-months to produce a working, if not very efficient, supercomputer version of the package. It is anticipated that, thereafter, the AIPS package will be used on one or more Cray machines for processing actual observations regardless of the outcome of the current plan. Experiments to improve the efficiency of the package will depend upon the prospects for this plan and/or upon the amount of use given to the preliminary version.

7.3.2 - Vectorizing Key Algorithms

During the past year NRAO personnel have been studying how to port NRAO's high performance vectorized algorithms from the AP-120B to other kinds of vector devices, both APs and supercomputers. In the course of this study several critical problems have been identified, for which optimum solutions are only partially known at present. Since it is extremely important to chose the computer which is appropriate for our needs, two of these problems as analyzed in some detail in Appendix A.

8. - COST SUMMARY

Table 8

(All figures in thousands of 1984 dollars)

Capital Costs

Computer Equipment	15,000
Physical Plant	580
Communications	800
Image Displays	540

	16,920

Incremental Annual Operational Costs

Computer Operations	2,100
Power	100
Communications	210

	2,410

Appendix A - CRITICAL ALGORITHMS

A.1 - The AIPS Pseudo Array Processor Concept

A number of AIPS tasks are designed to utilize a Floating Point Systems (FPS) AP-120B array processor when one is available. In order to use such tasks on machines which do not have an FPS AP, NRAO has developed the concept of a "pseudo-AP," in which a Fortran COMMON is used as the AP memory and Fortran or assembly language routines operate on data in this COMMON. These routines have exactly the same names, arguments, and functionality as the corresponding routines in the libraries for the AP-120B. There is only one version of each of the tasks which use the AP; the choice of whether it uses a true AP or the pseudo-AP is made by link editing it with the appropriate subroutine library. The "pseudo-AP" concept in AIPS actually amounts to defining a "virtual device interface" for vector processing.

The pseudo-AP memory space is called APCORE(), and is defined in the common block called /APFAKE/ which is declared in each subroutine of the pseudo-AP library. It is dimensioned elsewhere.

A.2 - Example 1 (APGRD4)

Two of the subroutines in the pseudo-AP library, APGRD4 and CLNSUB, contain concise examples of several of the data processing problems. They can be used as test cases to examine vector processing architectures. Listings of both of these subroutines are shown in Figures 6 and 7.

APGRD4 interpolates visibility samples onto a regular grid. The grid is required in order to allow the use of the Fast Fourier Transform (FFT) to compute the map. There are three main problems in vectorizing APGRD4 in various types of vector machines:

i) Data Dependent Addressing:

The integer subscripts JCX, JCY, and JG are computed in the outer DO 300 loop from the U-V coordinates of the visibility measurements. These subscripts are then used inside the DO 200 and DO 100 loops to access arrays in APCORE. Some APs have trouble passing addresses computed in floating point back to their addressing generators. This is generally not a problem in supercomputers.

ii) Short Vectors:

The values of variables JM and N, which are the loop limits of the DO 200 loop and the DO 100 loop, are generally of order 7 in AIPS applications. Many vector pipelines are not very efficient when processing vectors of length 7. A notable exception are the Cray computers. They will probably perform with at least 30% efficiency on this code on the first try because their vector pipelines have a low startup overhead. But the Cyber-205, a long-vector machine, usually works best for vectors longer than 50 elements. There are several possible approaches to improving the efficiency. The programmer can unroll the inner loop (i.e. write out the code 7 times with the indexing coded into it), and thereby eliminate some overhead at the cost of losing flexibility (the loop limit is hardcoded). A better solution for long-vector machines appears to be to permute the order of the three DO-loops so that the outer loop becomes the inner loop. This helps because JNVIS, its loop limit, is usually of order 500 or more. Such a modification of APGRD4 would allow a Cyber 205 to approach its theoretical performance. Indeed, the modification might also run better on Cray computers, allowing them to approach maximum performance as well.

iii) Vector Dependencies:

A pipelined processor can only overlap the beginning of processing one vector with the end of processing the previous one if the two vectors do not overlap in memory. If this rule is violated improper computations may result. This problem arises in APGRD4. The locations in memory where the two inner loops should perform their vector operations are computed from the visibility coordinates, as discussed in item 1 above. It is perfectly possible, indeed even likely, that successive visibilities will refer to the same cells in memory, so that accesses to APCORE() might violate the vector dependency rule. If the order of the DO-loops is permuted in order to improve performance on a Cyber 205 the dependency will appear in the inner loop! It appears that vector instructions can be used to detect the presence of dependencies in a particular vector of addresses and conditional vector instructions can be used to achieve the desired results. Further exploration of these ideas is needed.

A.2 - Example 2 (CLNSUB)

CLEAN is an iterative non-linear deconvolution algorithm which is used extensively in aperture synthesis radio mapping. Its purpose is to decompose an image into an ensemble of delta functions which, when convolved with the beam (the point source response of the interferometer array) will be equal to the original data. When the iterative decomposition process has converged the delta functions are reassembled into a new map in which the confusing sidelobe structures of the original "dirty" map are eliminated (see Figure 2). Radio astronomers refer to the new map as the "clean map." The most efficient version of CLEAN was designed by Barry Clark for the FPS AP-120B and is generally called the "Barry Clark CLEAN Algorithm." The heart of his algorithm is contained in a microcode routine called CLNSUB. CLNSUB subtracts one component from the current residual map and then finds the next component to be subtracted. It is called many thousands of times during the iterative decomposition process for a typical map. The AIPS pseudo-AP version of CLNSUB is shown below.

CLNSUB has two DO-loops. The second one, DO 200, is trivial for almost any vector processor: it merely finds the element of a vector which has the largest absolute value. The first loop, DO 100, appears to be an especially bad case for vectorizing because it contains IF-statements inside the loop (note that the outer loop of APGRD4 contains an IF, but it is only a minor problem in that case, especially if gather/scatter hardware is available). Most, if not all, current vectorizing compilers will refuse to consider vectorizing the DO 100 loop of CLNSUB. An obvious solution to this problem is to decompose the statements of the loop into a series of separate DO-loops, one for each statement. This means that a vectorizing compiler will be likely to compile each loop as a primitive vector operation. The IF-statements would execute vector operators which produce vectors of truth values (e.g. bit vectors on the Cyber 205) and subsequent vector operations would be conditioned on the truth of these vectors. The full ramifications of this transformation of the code have not yet been explored by NRAO's programmers.

```
      SUBROUTINE APGRD4 (JUV, JVIS, JWT, JGRID, JCONX, JCONY,
     *                           JNO2, JM, JLROW, JINC, JNVIS)
C-------------------------------------------------------------
C            JUV  = Location of (u,v) values in cells.
C            JVIS = Location of (complex) visibilities.
C            JWT  = Weight for data. Assumes any tapering
C                   has already been done.
C           JGRID = base address of gridded data.
C           JCONX = base address of X convolving fn.
C           JCONY = base address of Y convolving fn.
C            JNO2 = INT( (# cells used on a row) / 2 )
C              JM = number of rows kept in the AP.
C           JLROW = length of a row ( max. X).
C           JINC  = increment for UV,VIS and WT
C           JNVIS = number of visibility points to grid.
C-----------------------------------------------------------------
      INTEGER*2  N, INCR, HAF
      INTEGER*4  JUV, JVIS, JWT, JGRID, JCONX, JCONY, JCX, JCY,
     *           JA, JG, JJCX, JINC, JNVIS, JNO2, JM, JLROW,
     *           JJLOOP
      REAL*4     AIM, CWT, RE, X, XX, XWT, Y
      REAL*4     APCORE(1)
      COMMON /APFAKE/ APCORE
C-----------------------------------------------------------------
      IRND (XX) = IFIX (XX + SIGN (0.5, XX))
      JUV   = JUV + 1
      JVIS  = JVIS + 1
      JWT   = JWT + 1
      JGRID = JGRID + 1
      JCONX = JCONX + 1
      JCONY = JCONY + 1
      N     = JNO2 * 2 + 1
      HAF   = JLROW / 2 - JNO2
      INCR = 2 * JLROW - 2 * N


C                                        Loop over visibilities.
      DO 300 JJLOOP = 1,JNVIS
C                                        Check weight.
         XWT = APCORE(JWT)
         IF (XWT.LE.0.0) GO TO 300
C                                        Determine location.
         X = APCORE(JUV+1)
         Y = APCORE(JUV)
C                                        Deter. conv. fn loc.
         JCX = JCONX + IRND (10C. * (IRND (X) -
     *                                    X - 0.5)) + 100
         JCY = JCONY + IRND (100. * (IRND (Y) -
     *                                    Y - 0.5)) + 100
C                                        Determine grid loc.
         JG  = JGRID + 2 * (IRND (X) + HAF)
C                                        Save JCX
         JJCX = JCX
C                                        Get visibility
         RE  = APCORE(JVIS)   * XWT
         AIM = APCORE(JVIS+1) * XWT
C                                        Gridding loop
         DO 200 IY = 1,JM
            JCX  = JJCX
            RRE  = RE  * APCORE(JCY)
            AAIM = AIM * APCORE(JCY)
            DO 100 IX = 1,N
C                                        Sum to grid.
               APCORE(JG)   = APCORE(JG)   + APCORE(JCX) * RRE
               APCORE(JG+1) = APCORE(JG+1) + APCORE(JCX) * AAIM
C                                        Update pointers.
               JCX = JCX + 100
               JG  = JG  + 2
 100           CONTINUE
C                                        Update pointers.
            JCY = JCY + 100
            JG  = JG  + INCR
 200        CONTINUE
C                                        Update for next vis.
         JUV  = JUV  + JINC
         JVIS = JVIS + JINC
         JWT  = JWT  + JINC
 300     CONTINUE
      RETURN
      END
```

Fig. 6

```
      SUBROUTINE CLNSUB (JCOMP, JAMAP, JLMAP, JABEAM, JBBEAM)
C-----------------------------------------------------------------------
C   FORTRAN version of Barry Clark Microcode routine
C   CLNSUB does a CLEAN on the map points in the pseudo AP
C   using the portion of the beam in the pseudo AP.
C
C      JCOMP  Base address of the component vector:
C                  0 => intensity
C                  1 => X in cells
C                  2 => Y in cells
C                  3 => CLEAN loop gain (fractional)
C      JAMAP  Base address of the map stored as (loc,intensity)
C                  Loc = (2**14) * X + Y (cells)
C                           -1<X<NX, -1<Y<NY
C      JLMAP  Number of map points
C      JABEAM Base address of the piece of the beam.  The Y
C                  dimension varies the fastest: -BY<Y<BY,
C                  X varies slowest -1<X<BX
C      JBBEAM Base address of beam descriptor vector (BX,BY)
C-------------------------------------------------------------------
      INTEGER*4 JCOMP, JAMAP, JLMAP, JABEAM, JBBEAM,
     *          JAMAPO, LOOP, BEAMO, IBX, IBY, IDX, IDY
      REAL*4    XLOC, X, Y, DX, DY, XBEAM, SUBT, C1, C2,
     *          XCOMP, YCOMP, XMAX, AXMAX
      REAL*4    APCORE(1)
      COMMON /APFAKE/ APCORE
C                                       C1 = 2**(-14)
C                                       C2 = 2**14
      DATA C1, C2 /6.103515E-5, 16384.0/
C-----------------------------------------------------------------------
      JCOMP  = JCOMP + 1
      JAMAPO = JAMAP
      JAMAP  = JAMAPO + 1
      JBBEAM = JBBEAM + 1
      IF ( JLMAP.LE.0 ) GO TO 999
C                                       Get component to be CLEANed
      XCOMP = APCORE(JCOMP+1)
      YCOMP = APCORE(JCOMP+2)
      SUBT  = APCORE(JCOMP) * APCORE(JCOMP+3)
C                                       Get beam patch
      IBX = APCORE (JBBEAM)  + 0.1
      IBY = APCORE (JBBEAM+1)+ 0.1
C                                       Subtraction loop
      DO 100 LOOP = 1, JLMAP
C                                       Get next residual loc
         XLOC = APCORE(JAMAP)
         X    = XLOC * C1
         Y    = XLOC - IFIX (X) * C2
         DX   = X - XCOMP
         DY   = Y - YCOMP
         IDX  = DX + SIGN (0.1, DX)
         IDY  = DY + SIGN (0.1, DY)
C                                       Check if in beam patch.
         IF ((IABS(IDX).GE.IBX).OR.(IABS(IDY).GE.IBY)) GO TO 50
C                                       Get beam value
            IF (IDX.GE.0) BEAMO = IDY + IBY + IDX * (2 * IBY - 1)
            IF (IDX.LT.0) BEAMO = IBY - IDY - IDX * (2 * IBY - 1)
C                                       Subtract
            APCORE(JAMAP+1) = APCORE(JAMAP+1)
     *                                SUBT * APCORE(BEAMO+JABEAM)
C                                       Update pointer
 50         CONTINUE
         JAMAP = JAMAP + 2
 100     CONTINUE
C                                       Find largest mag. residual
      AXMAX = -10.0
      XMAX  = 0.0
      JAMAP = JAMAPO + 2
C                                       Loop
      DO 200 LOOP = 1, JLMAP
         IF (AXMAX.GE.ABS(APCORE(JAMAP))) GO TO 150
C                                       New maximum.
            AXMAX = ABS (APCORE(JAMAP))
            XMAX  = APCORE(JAMAP)
            XLOC  = APCORE(JAMAP-1)
 150        CONTINUE
         JAMAP = JAMAP + 2
 200     CONTINUE
C                                       Crunch location
      X = XLOC * C1
      Y = XLOC - IFIX (X) * C2
      APCORE(JCOMP)   = XMAX
      APCORE(JCOMP+1) = IFIX (X + SIGN (0.1, X))
      APCORE(JCOMP+2) = IFIX (Y + SIGN (0.1, Y))
 999  RETURN
      END
```

Fig. 7