National Radio Astronomy Observatory

COMPUTER ADVISORY GROUP DOCUMENTATION

SCIENTIFIC REQUIREMENTS FOR COMPUTER RESOURCES

August 1984

Table of Contents

1. - INTRODUCTION

This document outlines the scientific reasons for planning increased computing resources for the NRAO. Four sections describe the science to be done with four current or planned NRAO telescopes. The computing requirements to do that science have been estimated in some detail and form the basis for the plan to greatly increase the computing power available to NRAO.

The Very Large Array (VLA) is considered first and in greatest detail because it will almost certainly dominate NRAO's computing requirements for many years. It is the only telescope of the four which is currently operating and so the estimates of its scientific capabilities are much more certain than for the others. The discussion relating to the VLA has been taken directly from VLA Scientific Memorandum No. 150 "Astronomical Requirements for Future VLA Processing". Appendix A discusses a "canonical" VLA imaging task in terms of a conventional computer system and arrives at a conclusion about the computing power needed to support that task. Appendix B provides discussion of the conversion of scientific requirements into computing power needed to support the VLA. The appendices have been taken from VLA Computer Memorandum No. 168 "A Computer Plan for the VLA".

2. - ASTRONOMICAL REQUIREMENTS FOR FUTURE VLA PROCESSING

The purpose of this report is to estimate the long range computing power which the VLA will need in order to satisfactorily handle the data flow and computational power in the late 1980's and 1990's. The heart of this report is contained in Tables 2 and 3 where several of us (RDE,EBF,FNO, Pat Palmer and Jacqueline van Gorkam) have summarized the major VLA projects, their I/O and computational requirements, expected over the next ten years.

The interpretation of these requirements in terms of alternative computer configurations are given in the report 'A Computer Plan for the VLA', by R. Duquet, G. Hunt and R. Burns, VLA Computer Memorandum No. 168. This report will be referred to as DHB.

The general outline of this document is as follows:

1. Evaluation of the Present Situation
   - A. Computer systems for VLA data analysis
   - B. The major reduction and analysis tasks
   - C. Present capabilities
   - D. Present bottlenecks

2. Anticipated Projects at the VLA in Five years
   - A. Continuum projects
   - B. A digression about large field maps
   - C. Spectral line projects
   - D. Other projects at NRAO
   - E. Requirements for new hardware
   - F. Requirements for new software
   - G. Requirements for display

3. Estimation of Future Computer Demands

1. Evaluation of the Present System

The following section is, by and large, a summary of the discussion of the VLA computer resources as described by DHB. Here, we wish to stress the astronomical requirements and compare them with the existing systems.

A. Computer systems for VLA data analysis:

At present four computer systems handle the bulk of the VLA reductions and they are described in DHB. They are:

1) MODCOMP on-line system at the VLA which collects, correlates and stores the visibility data.
2) DEC-10 system at the VLA which calibrates and edits the visibility data.
3) Four mapping and image display systems running AIPS software presently handle most of the reductions and analyses from the point of calibrated visibility data to a final product. Three of these systems are run on a VAX 11-780 computer system.
4) PIPELINE, consisting of various PDP 11-series computers,

array processors and special purpose hardware, which will soon increase the map making and cleaning capacity at the VLA.

B. The major reduction and analysis tasks:

The approximate breakdown of the percentage of computing resources now devoted to the major VLA reduction tasks are given in Table 1.

TABLE 1

PROPORTION OF TASK USEAGE AT THE VLA

TASK	COMPUTER SYSTEM	PRESENT PROPORTION
On-line data collection	MODCOMP	not included
Visibility calibration	DEC-10	10%
Visibility I/O	DEC-10	7%
Mapping	VAX, PIPELINE	20% **
Deconvolution	VAX, PIPELINE	37% **
Selfcalibration	VAX	4% **
Tape to disk I/O	DEC-10, PIPELINE, VAX	7%
Displays of all kind	DEC-10, PIPELINE, VAX	9%
Map Analysis	VAX	6%

** heavy use of array processor not considered in percentage of cpu useage.

The table was derived from the task statistics in the VAX at the present time. The visibility calibration and I/O for the DEC-10 entry was calculated by assuming that its cpu power was about equal to that of one VAX 11-780 and that about 50% of the DEC-10 cpu time was used for data reduction. It is clear that the mapping and deconvolution take a majority of the computer resources and they will probably remain the most demanding tasks in the future. Both of these tasks and several others use the Fast Fourier Transform (FFT) as the crucial algorithm so a reasonable approximation to the VLA computer load would be to sum the rate of FFT's necessary to map, deconvolve and self-calibration a representative sample of observing projects.

C. Present capabilities:

In order to derive the present computational and data transfer capabilities for VLA data reduction, several bench mark tests were run on the VAX 11-780 and the PIPELINE systems. These tests included mapping a field 1024x1024 in size with 500,000 input visibility points; cleaning the resultant map with 3000 components and self-calibration of this visibility data base. From these and other tests, The present computational capabilities available at the VLA have been calculated by DHB in their Table 2. They find a total available computing power of 50 MIPS (millions of instructions per second) but a maximum useable computing power of about 15 MIPS with the present set of hardware. The efficiency of 30% is limited by the rate of data transfer to and from disk and cpu. In terms of the computational unit of a complex 2048 x 2048 FFT, which executes about 1.3 billion machine instructions, the 15 MIP computer power is equivalent to 500 FFT's in 12 hours.

Computer facilities outside of NRAO can help alleviate some of the computing load so the AIPS software in the VAX systems has been exported to many institutions. At the present time about 20 VAX's (4 with an array processor) are running AIPS about 15% of the time. The number of EXPORT systems and the average AIPS useage is expected to grow over the next five years. These total resources outside of NRAO are somewhat less than the NRAO resources and would not be able to handle the larger VLA problems. The outside systems will be desirable, especially in the latter stages of data analysis where the needed volume of data and the computational power is relatively low, and the advantage of performing the analysis in the home institution is large.

D. Present Bottlenecks:

The MODCOMP on-line system is nearly independent of the other VLA systems and its upgrading and improvement will be handled outside of the other NRAO/VLA computing systems. The MODCOMP system is now overloaded and minor improvements in the software and in the hardware are difficult to add. Plans for future development of this system are described in VLA Computer Memoranda No 166.

At the present time the DEC-10 is overloaded because all visibility data must pass through the system. When the PIPELINE begins routine reductions only the calibrator data will normally be processed in the DEC-10 and that system should be adequate for the calibration of VLA data.

It is in the area of mapping and deconvolution where the greatest bottleneck now occurs and will continue to occur in the future. This situation is largely the result of the increased use of the powerful deconvolution and self-calibration algorithms. These algorithms were largely unknown when the current VLA computer system was designed but they are now routinely used to improve the map quality by more than an order of magnitude. Many projects now require maps of size 2048, 4096, or larger, which are all but impossible to produce and clean. The PIPELINE, when fully operational, will be able to handle these larger maps, although not with the maximum desired throughput or flexibility.

The area of map analysis is presently software limited rather than computer limited; and may remain so in the forseeable future. Computer systems outside of NRAO can contribute signficantly to this facet of VLA data reduction because it requires much user interaction but does not require such large amounts of data storage and computing power. This analysis is most often handled by AIPS software on VAX computer systems and these systems are probably adequate to handle map analysis over the next five to ten years. At the present time, however, mapping, deconvolution and self-calibration monopolize the resources. As these tasks migrate to the PIPELINE and new generation NRAO computing systems, the existing systems (4 at NRAO and about 20 elsewhere) can adequately handle the map analysis and form a basis of evolution to the more powerful systems over the next 5 to 10 years.

An astronomer is forced, because of the present limitations, to map only that part of the primary beam which is of immediate interest. Often, the effects of strong sources outside the "interested" field of view distort the small maps and large amounts of computer resources and astronomer's time are wasted in trying to ascertain what is wrong with the data. A large field of view would have uncovered the strong sources and saved much computing time.

Perhaps even more importantly, this "tunnel vision" seriously reduces the chance of accidental discoveries of unusual or unsuspected radio emission outside of the main object. Such serendipitous discoveries have played a crucial role in the development of astronomy (eg. M. Harwit's book on Cosmic Discoveries). The VLA computer hardware and software should not exclude routine full field mapping.

2. Anticipated Projects within the next five years

A. Continuum projects:

The continuum projects and the observational parameters which determine their computational and I/O requirements are given in Table 2. These requirements are based on reasonable projects and represent those of moderate difficulty and completeness. These are not worst cases. Some of these projects can be enhanced by hardware improvements (more correlator channels, >35 km baselines and wider IF bandwidth). A detailed description of the assumed and derived values for each column is given at the end of Table 2.

Full Field Mapping:

These include objects or groups of objects which fill the entire primary beam. Some examples are individual large galaxies, clusters of galaxies or counts of background sources. Especially in the larger arrays the field of view at full bandwidth is limited by bandwidth smearing. To obtain maximum signal-to-noise over the full field it would be necessary to use the spectral line mode to subdivide the 50 MHz bandwidth into smaller channels which are combined into a single map where the correct (u-v) coordinates are calculated for each frequency band. The number of channels presently available is limited by the correlator so the entries in Table 2 correspond to less than the full 50 MHz. If the correlator is expanded to provide more channels at the maximum bandwidth the number of channels would be increased for those projects with an asterisk in column 8. Only a modest increase in the computer capacity results since all channels are still combined into a single map.

Observations of this type would normally be over six hours long in order to obtain adequate (u-v) coverage. The entry of 6 observations in a 12 hour period results from the requirement to subdivide the observation in the data reduction in order to correct for non-coplanar and corrections which rotate with the primary beam. The field of view entered in Tables 2A and 2B is the half-power primary beam width although some objects will require the full primary beam width.

Small Objects:

For small objects the area of interest is determined by the size of the object which is significantly less than the size of the primary beam. We then have two cases. First, if the object is weaker than the background sources, which is generally true at frequencies 5 GHz or less, the area of the primary beam must be mapped to correct for confusion. At 5 GHz the confusion most likely occurs from only one or two discrete sources in the field and in this case we have specified a low resolution, larger map in the table. At 75 and 327 MHz all observations will be severely confused.

Secondly, if the object is stronger than any confusing sources, only an area twice as large as the object need be mapped. This applies at frequencies 15 GHz or greater and at lower frequencies for all strong sources.

The entry of 12 observations per day is required to correct for the primary beam ellipticity, non-isoplanicity of the field and for the non-coplanar aperature. These problems are discussed in more detail in Section 2B.

Snapshots:

These are objects which are sufficiently bright and confined that signal-to-noise and (u-v) coverage is sufficient in a short observation and large stastistical samples can be analysed. At 1.5 and 5 GHz all fields will be confused so that the entire primary beam should be mapped, as described above. The eventual possibility of doing snapshot observations in spectral line mode to give maximum sensitivity has not been included. At the higher frequencies the largest area assumed is that unaffected by bandwidth smearing in which case the parameters scale with frequency and array; hence, the single entry in the table. Observations for which this assumption is not true are included under full field mapping.

Surveys by scanning the primary beam:

These are the limiting case of the snapshot observations in which the observational aim is to cover the largest possible area on the sky. The two examples cited are from actual proposals and are indicative of the range of parameters involved. With greater computer resources it is likely that this class of observation will become more common in the future.

Point sources:

Point source observations which are unaffected by confusion are listed here. Parameters are only given for the A configuration at 5 GHz since the processing load is relatively insensitive to frequency or array.

Solar:

These include observations of large images with variable structure. The sampling time constant is determined by the variability rather than (u-v) coverage. Possible extensions of the VLA hardware to provide shorter time constants have not been included.

Phased array:

Since the phased array results in only a few output channels, no significant data processing load results. It is included for completeness.

B. A digression about large field maps:

A major computational uncertainty in Tables 2, 3A and 3B is associated with projects which have a long integration time and a large field of view. Two complexities invalidate the use of the 2-D FFT for producing accurate maps from the (u-v) data. First, a non-circularly symmetric primary beam response means that the effective primary beam correction is a function of time since the primary beam rotates with respect to the sky because the antennas are on an alt-az mount. The time scale for such changes is several hours except when a source passes close to the zenith. This correction has been ignored in VLA reduction but it is significant for full field mapping in all configurations. It is believed that the limitiation of 100:1 in dynamic range in the C- and D-configurations is mainly produced by this non-circularity problem. Linear polarization maps are probably limited to 5% and circularly polarized maps to as much as 10%.

Assuming that the primary beam response is known, two correction methods are possible. The long integration can be broken in several short pieces (snap-shots), each reduced separately with its peculiar primary beam correction, and then the set summed to give the map associated with the entire integration. Alternatively, the primary beam correction (multiplicative in the map plane) can be applied as convolution to the (u-v) in the relevant time segment. This method, however, may cause some problems with the clean deconvolution.

A more serious problem is caused by the inadequacy of the 2-D FFT to produce undistorted maps of a large field of view from a non-planar aperture. Although the VLA is a nearly flat array, the aperture synthesized over several hours or longer is not planar in general. The phase error introduced by using the 2-D FFT is proportional to the product of th departure of the aperture from a plane (about equal to the length of the array) with the departure of the sky from a plane (equal to the distance-squared from the field center). For a given map size the phase error increases with wavelength and for the A-configuration at the VLA it is a serious problem at 1.4 GHz or less. Table 4 shows the size of the phase error, W, at several VLA frequencies and configurations and for the VLBA.

TABLE 4

Phase Errors Associated with the 2-D FFT

Frequency (GHz)	Configuration	Field of View (arcmin)	W (rad)
0.327	A	260.0	170
0.327	B	260.0	52
1.4	A	30.0	9
1.4	B	30.0	3
5.0	A	10.0	3
1.4	VLB	10.0	114
1.4	VLB	1.0	1
5.0	VLB	10.0	400
5.0	VLB	1.0	4

The effect is most serious in A-configuration and at the lower frequencies. The VLBA will be concerned with the W-term as well. Neglect of this phase term causes a point-source to appear 'U-shaped', the size of the U varying with the distance squared from the phase center. Sidelobes from this distorted source are not correctly removed using clean since the beam shape is not invariant with position.

There are several methods for dealing with the W-term. For a short period of time the synthetic aperture is planar and a true map of the entire field of view can be obtained using a 2-D FFT. The number of snap-shots needed is equal to about 2*W and each must be mapped and cleaned separately (some consolidation of cleaning and self calibration is probably possible). This solution is identical to one proposed for the non-circularity of the primary beam corrections. The VLBA, however, is not a planar array because of the curvature of the Earth so this option is not available.

A 3-D FFT can be used on the (u-v-w) data to form an (x-y-z) volume distribution. The width in the w-plane is also 2*W. A meaningful deconvolution solution must be constrained to lie on the celestial sphere in the (x-y-z) volume, something which the clean algorithm could handle with minor modification.

A third alternative is called mozaicing. Instead of making one large 2-D FFT map with distortions a whole set of maps, covering the field of view, but each sufficiently small to avoid the distortions, is made. The number of maps is about W*W. The main drawback of mozaicing, apart from the large number of maps, is that the sidelobe or alias responses of sources outside of the small map cannot be surpressed using clean. Perhaps the best way of elliminating the sidelobes is to subtract the sources directly from the (u-v-w) data and then remap. Subtraction from the observed (u-v) data is more accurate than subtraction from the gridded (u-v) data as done by the Clark version of Clean. For large fields of view which are dominated by a small number of strong, isolated sources, this method may be efficient.

The non-coplanar aperture problem has largely been ignored at the VLA because the hardware for generating sufficiently large maps is not at hand. At 327 MHz the W-term problem may be the most important limitation to accurate mapping. For 1.4 and 5.0 GHz in the A-configuration the problem is significant but not catastrophic.

In Tables 2, 3A and 3B we have assumed that any observation affected by the W-term or possible non-circularly symmetric primary beam response will be broken into 3, 6 or 12 snapshot observations over 12 hours with each segment reduced separately. It will take several years of experience to decide on the optimum procedure and the severity of the problems. We believe that this solution is realistic in estimating the computing and I/O power necessary. However, if we ever expect to routinely map the entire primary beam in the A and B configurations (serendipity again), correct maps must be made.

C. Spectral line projects:

A list of the spectral line projects and the observational parameters are tabulated in Tables 3A and 3B. As with the continuum, these are moderately difficult projects. Many spectral line projects are limited with the present correlator and Table 3B contains parameters for these projects assuming an enhanced correlator. A detailed description of the assumed and derived parameters follows Tables 3A and 3B.

This table includes all the major classes of spectal line mapping projects currently being pursued at the VLA. They are divided into groups with similar instrumental requirements. Because of the great range in possible parameters for spectral line observing we have not tried to cover all possible array combinations but tried to pick the typical bad but not extreme cases. In each case we have tried to set the specifications by asking what parameters would provide useful astronomical information if there were no limit imposed by the computer on channel number or pixel size; but given the resolution and sensitivity of the VLA. In making this analysis it also became clear that some modest extensions of the present VLA correlator would also provide useful additional capacity. Although these enhancements are not included in Table 3A, they have been include in Table 3B to give an indication of possible future expansions which should not be excluded by too modest a long-term computing plan.

Extragalactic Emission:

For these projects the velocity range is set by the dynamics of the galaxy or cluster of galaxies. The velocity and spatial resolution are mainly limited by brightness sensitivity. Two orthogonal polarizations are assumed to optimize signal-to-noise. In most of these cases the resolution and image size-scale with distance so that the numerical map size is the same.

Extragalactic absorption:
The relatively strong continuum source makes higher resolution possible, but the image size is now limited by the size of the continuum source.

Galactic Center:
This is separated from the other galactic projects because of the greater velocity range required.

Galactic absorption:
Again, resolution is not limited by brightness sensitivity so large numerical map sizes and channel numbers are possible. For OH and H2CO the full primary beam must be mapped to avoid confusion from HI emission. The velocity range used for NH3 is set to cover 3 transitions simultaneously. The channels between these lines need not be mapped.

Galactic emission:
Although the thermal sources are again limited in spatial and velocity resolution by the brightness sensitivity, the maser sources can be observed with the highest spatial and velocity resolution. The recombination lines include H, He and C. They can be observed in all VLA bands but 15 GHz is taken as typical.

Stars:
The regions of stimulated emission are much smaller for these masers. In some of these cases it may be necessary to use a self-calibration procedure which includes all the different channel maps in the model.

D. Additional Projects at NRAO:

VLBA:
Within five years the NRAO may be operating a ten-element VLB array and it is important to comment on its data reduction impact. The reduction and analysis of VLBA data will be virtually identical to that of the VLA so anticipating both VLA and VLB computer problems seems sensible. It is estimated that VLB reduction and analysis, from mapping to a "final" product will take about 25% of the computing power needed by the VLA as outlined in Tables 2 and 3A (See Chapter V, VLBA Report, May 1982). We assumed that the computer requirements will roughly scale with the number of correlators. It must be emphasized that this estimate is very tentative and will depend on the nature of the radio emission associated with milliarcsecond structure. We have also assumed that the map sizes for VLB objects will not be larger than those studied by the VLA. It is possible that a joint USA/Canada VLB array could consist of up to 19 elements, with an estimated load of 50% of the VLA.

The VLA and VLBA may operate as one large array of 37 antennas in the 1990's. While it is unlikely that all elements with spacings from 1 km to 3000 km would be used to make a high resolution, large-field map, it does seem appropriate to use the VLA with the 5 New Mexico antennas as a "super" VLA. This would increase the resolution limit but the large fields of view would remain. The cost in additional computing power would be significantly greater than for the 27-element VLA.

Millimeter array:

The projected millimeter array may add little extra computational demands although the I/O demands may be significant. Some new hardware technology and software development may be necessary for processing multi-feed synthesis data. Here, the most important point is to keep the future system as flexible as possible.

Proposed 75 MHz array:

The proposed 75 MHz extension of the VLA would significantly add to the computer load. Because of the density of sources in the 75 MHz beam, very large maps with large W-terms will be needed. A realistic estimate of the load would probably be equal to the 327 MHz entry in Table 2A as a minimum. When such a system is running, NRAO should have sufficient expertise in the problems discussed in Section 2B to handle the data. The large, expected ionospheric refraction and the resultant distortion over the field of view may necessitate advances in self-calibration to deal with this non-isoplanicity.

D. Requirements for new hardware:

A number of the projects listed in Tables 3B need a larger number of channels and bandwidth than is currently available. The increase in I/O and computing power with the correlator expansion is modest for the continuum projects, but increases the computer demands for the spectral line projects by a factor of 1.7. While we do not want ot argue that the propsed system must be able to handle the load given in the hypothetical Table 3B, it should be taken as an indication of possible future developments which ought not be designed out.

Other future hardware developments which may impact the computational requirements are:

Gating hardware for pulsar observations.
Fast sampling for Solar observations.
More observing flexibility.
VLA outstations, whether or not associated with the VLBA.

E. Requirements for new software:

New software techniques are continally being discovered and their impact on the long term computer planning is difficult to assess. For example, the self-calibration technique was unknown when the original VLA computer systems were designed. A description of some new techniques which might impact the VLA computer plans follows. Unfortunately, the computer and I/O power is now sufficiently limited so that the coding and analysis of new techniques are hampered. A listing of some of the algorithms currently used or under investigation is given in Appendix 4 of DHB.

Better weighting algorithms for (u-v) data before mapping should be investigated. The (u-v) tracks generally give a lumpy aperture coverage on both the small- and large-scale. By producing a smoother aperture with appropriate (u-v) weighting, the resultant maps would have lower sidelobes and better signal-to-noise.

Deconvolution is the most computer intensive task for VLA reductions. Algebraic- and maximum entropy-type algorithms have been investigated but they can be an order of magnitude more expensive than the CLEAN algorithm. The VLA resources at the present time are not sufficient to properly analyse and test the more sophisticated algorithms. It is unlikely that a faster algorithm than clean will be found. However, more investigations for optimal deconvolution of extended sources is needed. Subtraction of clean components from the ungridded (u-v-w) data is also needed for wide-field mapping problems.

As yet no serious attempts have been made to estimate the reliability of the deconvolved maps. Such error estimation algorithms are now known but all have required too much computation time to be investigated.

Self-calibration techniques have extended the dynamic range of the VLA maps from 100:1 to over 10,000:1. The algorithm is not particularly expensive, compared with clean, and in almost all cases it need only be made on one channel in a set of continuum or spectral line data. More sophisticated time filtering and parameter fitting capabilities are needed to self-calibrate weak sources and for fields at low frequency which may be severely distorted by ionospheric refraction.

The non-circularity of several corrections associated with the primary beam and the W-term distort maps made from long integrations on extended sources. These software problems have already been discussion in Section 2B. The relative costs of the several alternative are not as yet determined.

Software associated with map analysis after good quality images have been produced is open ended. The two computations which may be somewhat computer and I/O intensive are; 1) the profile analysis of a set of spectral line maps and 2) subtraction of the continuum radio emission from the line maps directly from the (u-v) data. Such analyses can be generally handled on AIPS computing systems now available at NRAO and many other institutions. It is likely that when much of the mapping, deconvolution and self-calibration is moved to the next generation NRAO computer, AIPS on a VAX-type computers (perhaps with upgraded array processors) will be able to handle map analysis and display anticipated from Tables 2 and 3.

We strongly emphasize the uncertainty in attempting to predict the important advances in new techniques of VLA reduction and analysis. It does seem likely that the new algorithm will be costly.

F. Requirements for display:

The main interface between the observer and the computer is by a display. Data calibration and editing, and map analysis are now somewhat hampered by the lack of creative display software. New technologies (e.g. vector graphics, see SPG memo 11 by J. Torson), greater display power, intelligent combination of graphics and reduction software would improve thoughput by aiding the observer in detecting bad data and in deciding the best course to follow in the subsequent reductions. Although these requirements should not impact the computation and I/O capabilities outlines in Tables 2 and 3 they will require additional display hardware.

3. Estimation of Future Computer Demands

The estimation of the future computer demands from the astronomical point of view comes directly from Tables 2 and 3; specifically from the average and rate of FFT's needed to process the data. The explanation of these tables gives the assumptions that were made in obtaining these parameters. The summary of demands is

TABLE 5

VLA COMPUTER DEMANDS

Type	2-D FFT's in 12 h	Ratio with Current Power
Continuum	3698	7.0
Continuum (enh)	3698	7.0
Spectral Line	8713	17.4
Spectral Line (enh)	13368	26.7

This estimate for the computing load is more than ten times than now currently available at the NRAO.

A detailed look at the computing and I/O demands in terms of computer configurations are analyzed in the report by DHB.

Finally, it should not go unnoticed that the problems described in this memorandum, and the resources needed to handle them, are such that this facility would be able to make a major impact on image analysis for all areas of astronomy.

TABLE 2

COL EXPLANATION

1...Class of project:

2...Number of observations in 12 hours to be separately reduced:
For long integrations on one field, mapping, and cleaning must be done on short segments of data in order to avoid non-coplanar aperture effects and to correct for non-circularly symmetric primary beam response. See discussion in Section 2B. For other projects this entry gives the number of separate fields.

3...Number of polarizations in input data:
2=only parallel polarizations; 4=all polarizations.
3=Average of the above two options.

4...Field of view in arcminutes:
Determined by either astronomical requirements or by the primary beam.

5...Angular resolution in arcseconds:
Determined either by astronomical requirements (such as brightness sensitivity) or by the VLA maximum baseline.

6...Number of pixels on map side:
3 (points per beam) * 60 * (field size) / (resolution) rounded to nearest hundred.

7...Sampling time required for less than 10% degradation:
277 * (resolution) / (field size); 60 sec maximum

8...Number of channels required for less than 10% bandwidth degradation:
2 * Number of pixels / 50 / freq(GHz), when limited by present correlator. Bandwidth is also less than 50 MHz. Both AC and BD IF's are assumed.

9...Percentage of fields requiring self-calibration:

10...Percentage of fields requiring deconvolution (clean):

11...Repetition factor:
Number of times data is mapped and cleaned before obtaining a map free of errors

12...Number of input words obtained in 12 hours:
Input = Pol * Chnls * 351 * 2 * 43200 / Sample time
Assumes two words (16 bits each) per input datum

13...Number of 2-D FFTs for mapping, cleaning and self-calibration:
NFFT = Nobs * [(Pol/2 + (Pol-1) * 12 * DC% / 100) * Rep
+ 30 * SC% / 100]
Discussion of the equation is given by DHB

14...Percentage of anticipated observing time in future:
This percentage is based on the current observing statistics and it is modified by the anticipated effects of instrumental and computer improvements.

15...Average number of equivalent 2048x2048 FFT's needed in 12 hours:
NFFTAVG = OBS TIME% / 100 * NFFT * Pix * Pix
* Ln(Pix) / 32,000,000

(1) Class of Project	(2) No of obs. (in 12hr)	(3) Pol.	(4) Field 'arc	(5) Resoln "arc	(6) Pixels	(7) Sample time sec	(8) Chnls	(9) Self -cal %	(10) Deconvo -lution %	(11) Repeat factor	(12) Input words	(13) No. of 2-D FFTs	(14) obser time %	(15) Average equivalent 2048 FFTs
Full field mapping														
A array														
1.4 GHz	6	4	30.0	1.0	5400	9	32 *	75	100	2	4.2E+08	591	4	185
5 GHz	6	4	10.0	0.3	6000	8	32 *	75	100	2	4.7E+08	591	5	289
15, 22 GHz	6	4	3.0	0.1	5400	9	12	75	100	2	1.6E+08	591	2	93
B array														
1.4 GHz	6	4	30.0	3.3	1600	30	32 *	75	100	2	1.3E+08	591	5	17
5 GHz	6	4	10.0	1.0	1800	27	12	75	100	2	5.3E+07	591	3	13
15, 22 GHz	6	4	3.0	0.3	1600	30	4	75	100	2	1.6E+07	591	2	7
C,D array														
1.4 GHz	6	4	30.0	11.0	500	60	16	75	100	2	3.2E+07	591	2	1
5 GHz	6	4	10.0	3.3	600	60	4	75	100	2	8.1E+06	591	5	2
15, 22 GHz	6	4	3.0	1.1	500	60	2	75	100	2	4.0E+06	591	3	1
Small objects														
Weak sources, 75 and 327 MHz	12	2	260.0 c	5.0	9400	5	2	100	100	2	2.3E+07	672	6	1019
1.4 GHz	6	4	30.0 c	1.0	5400	9	2	50	100	2	2.6E+07	546	6	257
5 GHz	6	4	10.0 c	1.2	1500	33	2	50	100	2	7.3E+06	546	5	14
15,22 GHz	6	4	0.5 a	0.3	300 a	60	2	50	100	2	4.0E+06	546	4	0
Strong sources, all freq.	6	4	0.5 a	0.3	300 a	60	2	100	100	2	4.0E+06	636	5	1
Snapshots														
Weak sources, 1.4 GHz	100	3	30.0 c	1.0	5400	9	2	50	100	1	2.0E+07	4050	4	1269
5 GHz	100	3	10.0 c	0.6	3000	17	2	50	100	1	1.1E+07	4050	5	456
15,22 GHz	100	3	0.5 a	0.3	300 a	60	2	50	100	1	3.0E+06	4050	4	3
Strong sources, all freq.	200	4	0.5 a	0.3	300 a	60	2	100	100	1	4.0E+06	13600	9	20
Survey by scanning primary beam														
B array, 1.4 Ghz, 1 obs/min	720	2	20.0	3.3	1100	46	2	20	100	1	2.7E+06	13680	1	36
C array, 1.4 GHz, 6 obs/min	4320	2	30.0	11.0	500	60	2	0	20	1	2.0E+06	14688	2	14
Point sources														
Astrometry (A array, 5 GHz)	100	2	0.1	0.3	100	60	2	0	0	2	2.0E+06	200	3	0
Monitoring, spectra	100	4	0.1	0.3	100	60	2	50	0	2	4.0E+06	1900	2	0
Flare stars	10	4	0.1	0.3	100	3 b	2	0	0	2	8.1E+07	40	1	0
Detections	24	2	0.1	0.3	100	60	2	0	0	2	2.0E+06	48	7	0
Solar														
Quiet	2	4	30.0	10.0	500	60	2	0	100	2	4.0E+06	152	1	0
Active	20	4	30.0	10.0	500	3 b	2	0	100	2	8.1E+07	1520	3	2
Phased array	0	2					1	100	0				1	
													---	------
													100	3698 FFTs/12hr

(a) Bandwidth limited field: scales with frequency and array so that number of pixels is constant.

(b) Averaging time determined by variability.

(c) Field of view must be full primary beam to remove confusing sources.

(*) Limited by the present correlator.

TABLE 3A and 3B

COL EXPLANATION

1...Class of project:
2...Frequency in GHz:
3...Number of observations in 12 hours to be separately reduced:
For long integrations on one field, mapping, cleaning and selfcalibration must be done on short segments of data in order to avoid non-coplanar aperture effects and to correct for non-circularly symmetric primary beam response. See discussion in section 2B. For other projects this entry gives the number of separate fields.
4...Number of polarizations in input data:
2=both parallel polarizations; 4=all polarizations.
5...Field of view in arcminutes:
Determined either by astronomical requirements or by the primary beam.
6...Angular resolution in arcseconds:
Determined either by astronomical requirements (such as brightness sensitivity) or by the VLA maximum baseline.
7...Number of pixels on map side:
3 (points per beam) * 60 * (field size) / (resolution) rounded to nearest 100.
8...Sampling time required for less than 10% degradation:
277 * (resolution) / (field size); 60 sec maximum
9...Velocity range required in km/s:
Determined by astronomical requirements
10...Maximum velocity resolution in km/s:
Determined by astronomical requirements and signal/noise
11...Number of channels:
Nch = 1.3 * (vel range) * (vel resol), unless the correlator specifications are exceeded (*) in which case the number of channels in Table 2A has been set to the maximum possible. This may result in a poor compromise between channels, bandwidth, polarizations and interferometer pairs, however, the computation load will be reasonable. A 30% range is included for baseline determination.
12...Percentage of channels for which separate beams are needed:
%Beams = Pixels * Vel res / 1800. Assumes less than 10% error in beam location at the edge of the map.
13...Percentage of fields requiring self-calibration:
14...Percentage of fields requiring deconvolution (clean):
15...Is subtraction of component from visibility data required?
16...Number of input words obtained in 12 hours:
Input = Pol * Chnls * 351 * 2 * 43200 / Sample time
Assumes two words (16 bits each) per input datum
17...Number of 2-D FFTs:
NFFT = Nobs * [(Pol-1) * 1.3 * Nch * (1 + BEAMS% / 100 + 10 * DC% / 100) + 15 * SC% / 100]
Assumes 10% of the channels are reduced three times as a repetition factor. Discussion of equation given by DHB.
18...Percentage of anticipated observing time in future:
19...Average number of equivalent 2048x2048 FFT's needed in 12 hours:
NFFTAVG = OB% /100 * NFFT * Pix * Pix * Ln(Pix) / 32,000,000

TABLE 3A

ASTRONOMICAL REQUIREMENTS - SPECTRAL LINE

(1) Class of Project	(2) Freq. GHz	(3) No. of obs (in 12hr)	(4) Pol.	(5) Field 'arc	(6) Space Resoln. "arc	(7) Pixels	(8) Sample time sec	(9) Range km/s	(10) Velocity Resoln. km/s	(11) Chnls	(12) No. beams %	(13) Self cal. %	(14) Deconvolution %	(15) uv sub.	(16) Input words	(17) No. of 2-D FFTs	(18) Observ time %	(19) Average equivalent 2048 FFTs
Extragalactic emission																		
HI clusters	1.4	1	2	30	5	1100	46	5000	20	16 ‡	12	0	20	yes	2.1E+07	65	6	1
HI galaxies	1.4	2	2	30 c	10 c	500	60	1000	10	64 ‡	3	0	75	yes	6.5E+07	1419	12	8
ISM (individ. gal	1.4	3	2	30 a	1	5400	9	500	5	130	15	0	50		8.5E+08	3118	6	1465
recomb. line	15	1	2	10 a	10	200	60	1000	10	8 ‡	1	100	50	yes	8.1E+06	78	1	0
NH3	23	1	2	2 a	5	100	60	500	2	8 ‡	0	0	10		8.1E+06	21	1	0
H2O masers	22	3	4	2 a	0.06 b	6000	8	500	2	4 ‡	7	0	0		5.8E+07	50	1	5
OH masers	1.6	3	4	30 a	1 b	5400	9	500	2	64 ‡	6	0	0		8.4E+08	794	1	62
Extragalactic absorption																		
HI	1.4	3	2	2	1 b	400	60	1000	10	64 ‡	2	100	50	yes	6.5E+07	1548	2	1
OH	1.6	2	4	2	1 b	400	60	500	10	65	2	100	50	yes	1.3E+08	3083	3	3
H2CO	5	2	2	2	0.3 b	1200	42	500	10	65	7	100	50	yes	9.5E+07	1055	2	7
Galactic center																		
NH3	23	2	2	2 a	5	100	60	400	1	16 ‡	0	100	50		1.6E+07	280	3	0
H2CO	5	1	2	9 a	0.3 b	5400	9	400	1	64 ‡	3	100	50		4.2E+08	517	1	40
HI, OH	1.4	1	2	30 a	1 b	5400	9	400	1	256 ‡	3	100	50		1.7E+09	2022	1	158
recomb. line	15	1	2	9	1	1600	31	1000	4	8 ‡	4	100	50		1.6E+07	78	1	0
Galactic absorption																		
HI	1.4	3	4	30 ad	1 b	5400	9	100	1	130	3	100	100		1.7E+09	16822	2	2635
OH	1.6	3	4	10	1 b	1800	28	100	1	130	1	100	100		5.7E+08	16791	2	255
H2CO	5	3	2	10	0.3 b	6000	8	100	1	130	3	100	100		9.5E+08	5639	2	1104
NH3	22	1	2	2 a	0.06 b	6000	8	3.2 e	0.01 e	128 ‡	0	100	100		9.3E+08	1845	4	722
Galactic emission																		
HI	1.4	2	2	10	5	400	60	200	5	52	1		50		5.3E+07	813	2	0
NH3	22	1	2	2 a	5	100	60	3.2 e	0.01 e	128 ‡	0	100	10		1.3E+08	348	12	0
recomb. lines	15	2	2	4 a	5	100	60	150	3	65	0	100	50	yes	6.6E+07	1044	11	0
OH masers	1.6	12	4	10	1	1800	28	100	1	130	1	100	50	yes	5.7E+08	36745	7	1952
H2O masers	22	2	4	2	0.06	6000	8	100	0.5	32 ‡	2	100	5	yes	4.7E+08	409	6	240
Stars																		
OH masers	1.6	12	4	0.2	1 b	100	60	60	0.5	156	0	100	50	yes	3.2E+08	43987	6	4
H2O masers	22	12	4	0.2	0.06 b	600	60	60	0.5	32 ‡	0	100	50	yes	6.5E+07	9168	3	20
SiO masers	44	12	4	0.2	0.03 b	1200	42	60	0.5	16 ‡	0	100	50	yes	4.7E+07	4675	2	30
																		----- 8713 FFTs/12h

(a) Limited by primary beamwidth.

(b) Limited by maximum VLA resolution.

(c) Scales with distance of object or frequency so that number of pixels is constant.

(d) Full primary beam needed for HI absorption because emission must be measured over the whole primary beam.

(e) Units are MHz. Specified to cover line structure.

(‡) Limited by specification of present correlator.

TABLE 3B ASTRONOMICAL REQUIREMENTS - SPECTRAL LINE (Expanded correlator)

(1) Class of Project	(2) Freq. GHz	(3) No. of obs (in 12hr)	(4) Pol.	(5) Field 'arc	(6) Space Resoln. "arc	(7) Pixels	(8) Sample time sec	(9) Range km/s	(10) Velocity Resoln. km/s	(11) Chnls	(12) No. beams %	(13) Self cal. %	(14) Deconvolution %	(15) uv sub.	(16) Input words	(17) No. of 2-D FFTs	(18) Observ time %	(19) Average equivalent 2048 FFTs
Extragalactic emission																		
HI clusters	1.4	1	2	30	5	1100	46	5000	20	325 *	12	0	20	yes	4.3E+08	1319	6	21
HI galaxies	1.4	2	2	30 c	10 c	500	60	1000	10	130 *	3	0	75	yes	1.3E+08	2882	12	17
ISM (individ. gal	1.4	3	2	30 a	1	5400	9	500	5	130	15	0	50		8.5E+08	3118	6	1465
recomb. line	15	1	2	10 a	10	200	60	1000	10	130 *	1	100	50	yes	1.3E+08	1031	1	0
NH3	23	1	2	2 a	5	100	60	500	2	325 *	0	0	10		3.3E+08	845	1	0
H2O masers	22	3	4	2 a	0.06 b	6000	8	500	2	325 *	7	0	0		4.7E+09	4056	1	397
OH masers	1.6	3	4	30 a	1 b	5400	9	500	2	325 *	6	0	0		4.3E+09	4031	1	316
Extragalactic absorption																		
HI	1.4	3	2	2	1 b	400	60	1000	10	130 *	2	100	50	yes	1.3E+08	3098	2	2
OH	1.6	2	4	2	1 b	400	60	500	10	65	2	100	50	yes	1.3E+08	3083	3	3
H2CO	5	2	2	2	0.3 b	1200	42	500	10	65	7	100	50	yes	9.5E+07	1055	2	7
Galactic center																		
NH3	23	2	2	2 a	5	100	60	400	1	520 *	0	100	50		5.3E+08	8143	3	0
H2CO	5	1	2	9 a	0.3 b	5400	9	400	1	520 *	3	100	50		3.4E+09	4091	1	320
HI, OH	1.4	1	2	30 a	1 b	5400	9	400	1	520 *	3	100	50		3.4E+09	4091	1	320
recomb. line	15	1	2	9	1	1600	31	1000	4	325 *	4	100	50		6.4E+08	2565	1	15
Galactic absorption																		
HI	1.4	3	4	30 ad	1 b	5400	9	100	1	130	3	100	100		1.7E+09	16822	2	2635
OH	1.6	3	4	10	1 b	1800	28	100	1	130	1	100	100		5.7E+08	16791	2	255
H2CO	5	3	2	10	0.3 b	6000	8	100	1	130	3	100	100		9.5E+08	5639	2	1104
NH3	22	1	2	2 a	0.06 b	6000	8	3.2 e	0.01 e	416 *	0	100	100		3.0E+09	5964	4	2335
Galactic emission																		
HI	1.4	2	2	10	5	400	60	200	5	52	1		50		5.3E+07	813	2	0
NH3	22	1	2	2 a	5	100	60	3.2 e	0.01 e	416 *	0	100	10		4.2E+08	1097	12	0
recomb. lines	15	2	2	4 a	5	100	60	150	3	65	0	100	50	yes	6.6E+07	1044	11	0
OH masers	1.6	12	4	10	1	1800	28	100	1	130	1	100	50	yes	5.7E+08	36745	7	1952
H2O masers	22	2	4	2	0.06	6000	8	100	0.5	260 *	2	100	5	yes	3.8E+09	3106	6	1824
Stars																		
OH masers	1.6	12	4	0.2	1 b	100	60	60	0.5	156	0	100	50	yes	3.2E+08	43987	6	4
H2O masers	22	12	4	0.2	0.06 b	600	60	60	0.5	156 *	0	100	50	yes	3.2E+08	43997	3	95
SiO masers	44	12	4	0.2	0.03 b	1200	42	60	0.5	156 *	0	100	50	yes	4.6E+08	44009	2	281

																		13368 FFTs/12h

(a) Limited by primary beamwidth.

(b) Limited by maximum VLA resolution.

(c) Scales with distance of object or frequency so that number of pixels is constant.

(d) Full primary beam needed for HI absorption because emission must be measured over the whole primary beam.

(e) Units are MHz. Specified to cover line structure.

3. - COMPUTING FOR THE VERY LONG BASELINE ARRAY

There has been considerable discussion lately of the need for a very large computing facility for the reduction of VLA data. The VLA requirements have been specified by VLA Scientific Memorandum No. 150 "Astronomical Requirements for Future VLA Processing" (hereafter referred to as Memo 150). It concludes that, in order to support several classes of computer intensive projects, a computing capability in the supercomputer class is required. In support of this conclusion, tables are presented that show the computer needs and the projected fraction of the total available observing time for the major classes of observations that would be done on the VLA in the absence of limitations imposed by the post-processing computers. If a major computing facility is acquired by NRAO, it presumably would be used to reduce data from all NRAO instruments including the Very Long Baseline Array (VLBA).

This memo is an attempt to specify the computing needs of the VLBA in a manner similar to that used by Memo 150. Its purpose is to show what science can and cannot be done with the post-processing computers specified in the VLBA proposal and to determine to what extent the needs of the VLBA help justify the acquisition of the large computing facility. The breakdown of the projected computing needs by class of observations is given in Tables 1 and 2 for continuum and spectral line, respectively. A description of the contents of the tables is given below. The numbers in the tables represent our best guesses at this time. But it must be kept in mind that they are guesses. Even the VLA numbers in Memo 150 are uncertain by large amounts despite the fact that the VLA has been in operation for several years. The actual scientific emphasis, not to mention the processing techniques, in use in 1989 when the VLBA is completed, may differ considerably from our current estimates.

The tables only address the problem of mapping the data. There is likely to be a large computing load for the VLBA associated with fringe fitting. It is possible that, with careful use of calibrators to fix the delay and rate of each antenna, fringe fitting will not need to be done on many sources. Conversly, fringe fitting can be improved if it is done globally with a good input model. Since the model is likely to be based on a map made with data from a preliminary fit, two fits may be required for a significant fraction of the data. Our current estimate is that fringe fitting will require the equivalent of one VAX 11/780 plus AP.

The tables specify a mix of observing that would occupy the VLBA for about 83 percent of the time. No attempt has been made to specify the full 100 percent because time will be needed for maintenance and for projects not in the table whose computer needs are unknown at this time (eg. solar).

Based on the results of VLA Computer Memorandum No. 168, "A Computer Plan for the VLA", a 1 Mflop computer could calculate about 144 2048 by

2048 FFT's per 12 hours (the units of compute power used in Memo 150 and in the last columns of the tables) without considering the inevitable inefficiencies due to I/O and overhead. The achieved compute rate is likely to be about half that. A VAX plus AP configured like the current AIPS machines is capable of about 1.5 Mflops. The needs specified in the tables are about 142 of these units so the three VAX 11/780 plus AP's that are specified in the VLBA proposal (in addition to the one needed for fringe fitting) should provide enough compute power to deal with the projects specified plus a few others not in the tables. Note that the VAX plus AP is used here as a generic unit of compute power to aid in cost estimates. By the time the computers are purchased (1987 or later), newer models will almost certainly be more attractive.

It must be emphasized that both the computing needs of the array and the capabilities of any given machine combination (especially in 1988 technology) are very poorly known. Changes in some of the parameters in the tables, or the inclusion of some known, severe cases, can drive the computing needs to totally unreasonable levels. For example, mapping HII region water masers by the brute force method (Single large X, Y, V cube as opposed to many small "cubicles") involves making about 400 maps, each 30,000 pixels on a side from a data set containing about 5500 million words. If about 20 such observations are made per year (HII regions + proper motions), the required number of equivalent 2048 2D FFT's per 12 hr. is about 130,000! And that is for a 3 arc second source. The masers in Orion are spread over 30 arc seconds. Clearly there are cases that cannot be done by brute force methods.

The tables are also restricted to projects involving primarily the VLBA. There are projects that involve the use of the VLBA with the VLA that will have severe computing requirements. The simplest cases, and perhaps the most severe, are those that involve the use of the Pie Town antenna of the VLBA to double the resolution of the VLA. The needs of these observations can be determined by doubling the size of the maps used in A array observations specified by Memo 150. The need to remove confusion will still be there (fringe rate and delay offsets make confusion unlikely to be a problem on the VLBA itself) so large fields will be needed. Doubling the size of the maps increases the computing load by more like a factor of 5 so the needs specified by Memo 150 will rise sharply. For this reason, and not so much for the VLBA as a stand alone instrument, the construction of the VLBA will increase the computer needs of NRAO. This increase is not supported within the VLBA budget.

Table I gives the projected computing needs for continuum observations. The meanings of the items in the table, by column number are:

1.) Class of object.
2.) The number of observations per 12 hour period.
3.) The number of polarizations.
4.) The field of view over which the data should not be degraded given in milli-arcseconds.
5.) The resolution in milli-arcseconds.
6.) The number of pixels per side of a map of the field. This is 3 * (field of view) / (resolution)
7.) The maximum sample time for less than 10% degradation.
    277 * 60 * (resolution) / (field size)
    This is the equation is the same as in Memo 150. Maximum integration time is 60 sec. Note that this allows the longest baselines to go through about .6 of a fringe and may be too long, especially when self cal is used.
8.) The number of channels required for less than 10% bandwidth degradation:
    (pixels) * (bandwidth(MHz)) / freq(MHz) / 2.5
    This is the same equation as in Memo 150 except that only 1 IF pair is assumed and no upper limit on the number of channels is assumed.
9.) The number of hybrid mapping loops used.
10.) The percent of time that a deconvolution algorithm is used.
11.) The number of tries it will take to make a final map.
12.) The number of input words obtained in 12 hours.
    (polarizations) * (channels) * (baselines) * (43200/sample time) * 3
    This assumes 3 words per datum (real, imag, and weight). Memo 150 assumed 2 words per datum.
13.) The number of 2-D FFTs needed for mapping, cleaning, and self-calibration.
    NFFT = Nobs * [(pol/2 + (pol-1)*12*decon%/100) * repeat + 13*Nhyb]
    The equation is very similar to that used by Memo 150 except that the number of hybrid loops is included.
14.) The amount of observing time that will be allocated to the class of object per year. The time is specified in terms of the equivalent number of 12 hr. observing runs. These numbers are wild guesses at this time.
15.) Average equivalent 2048 FFTs in 12 hr.
    EFFT = NFFT * (percent obs. time)/100 * (pixels**2)*log(pixels) / (2048**2)*log(2048)

TABLE I

(1) Class of Project	(2) No of obs. (in 12hr)	(3) Pol.	(4) Field mas	(5) Resoln mas	(6) Pixels	(7) Sample time sec	(8) Chnls	(9) Hyb loops	(10) Deconvo -lution %	(11) Repeat factor	(12) Input words	(13) No. of 2-D FFTs	(14) No. of 12 hr /yr	(15) Avg. equiv 2048
Monitoring Observations														
Superluminals	3	4	60.	0.5	360	30.0	1	10	100	2	0.78E+06	618	50.	1.
Other extragalactic	3	4	60.	0.5	360	30.0	1	10	100	2	0.78E+06	618	50.	1.
Galactic sources	3	4	200.	2.0	300	30.0	3	10	100	2	0.23E+07	618	20.	0.
Extragalactic Source Structure														
Compact cores	5	4	50.	1.0	150	30.0	1	10	100	2	0.78E+06	1030	50.	0.
Inner jets	2	4	400.	1.0	1200	30.0	3	10	100	2	0.23E+07	412	50.	9.
Weak Sources	1	2	200.	1.0	600	30.0	2	0	100	4	0.78E+06	52	50.	0.
High Dyn. Range (19 s	1	4	500.	1.0	1500	30.0	4	20	100	3	0.12E+08	374	50.	13.
Hot Spots (+VLA)	2	4	2000.	3.0	2000	24.9	18	10	100	2	0.34E+08	412	20.	11.
Galactic Objects														
SS433 etc.	2	4	400.	1.0	1200	30.0	3	10	100	2	0.23E+07	412	20.	4.
Astrometry/Geodesy														
Plate Motions	30	2	10.	0.5	60	30.0	1	1	25	2	0.57E+06	630	8.	0.
Astrometry	30	2	20.	0.5	120	30.0	1	1	25	2	0.39E+06	630	20.	0.
Proper Motions	50	2	20.	0.5	120	30.0	1	1	25	2	0.39E+06	1050	40.	0.
Pulsars	30	2	30.	0.5	180	30.0	1	1	25	2	0.39E+06	630	40.	0.
Array Calibration	30	2	20.	0.5	120	30.0	1	1	25	2	0.39E+06	630	20.	0.

Percent of year used: 67.0
Total effective FFTs: 39.8

The parameters for spectral line observations are given in Table II The meanings of the columns are:

1.) Class of object.
2.) The number of observations per 12 hour period.
3.) The number of polarizations or fields. When a source is assumed to be mapped using many small fields of view, this number is the product of the number of fields and the number of polarizations.
4.) The field of view over which the data should not be degraded given in milli-arcseconds.
5.) The resolution in milli-arcseconds.
6.) The number of pixels per side of a map of the field. This is 3 * (field of view) / (resolution)
7.) The maximum sample time for less than 10% degradation. 277 * 60 * (resolution) / (field size) This is the equation is the same as in Memo 150. Maximum integration time is 60 sec.
8.) The velocity resolution in km/sec.
9.) The velocity range to be covered in km/sec
10.) The number of channels = 1.3 * range/resolution.
11.) The number of separate dirty beams needed to avoid errors at the edge of a map by greater than 10%.
    Nbeam = pixels * vres / 1800.
    This is the same equation as in Memo 150.
12.) The number of hybrid mapping loops used.
13.) The percent of time that a deconvolution algorithm is used.
14.) The number of input words obtained in 12 hours.
    (polarizations) * (channels) * (baselines) * (43200/sample time) * 3
    This assumes 3 words per datum (real, imag, and weight).
    Memo 150 assumed 2 words per datum.
15.) The number of 2-D FFTs needed for mapping, cleaning, and self-calibration.
    NFFT = Nobs * [(pol-1) * 1.3 * chans * (1+Nbeams/100 + 10*Pdecon/100) + 15 * nhyb / 100]
    The equation is very similar to that used by Memo 150 except that the number of hybrid loops is included.
16.) The amount of observing time that will be allocated to the class of object per year. The time is specified in terms of the equivalent number of 12 hr. observing runs. These numbers are wild guesses at this time.
17.) Average equivalent 2048 FFTs in 12 hr.
    EFFT = NFFT * (percent obs. time)/100 * (pixels**2)*log(pixels) / (2048**2)*log(2048)

TABLE II

(1) Class of Project	(2) No of obs. /12hr	(3) Pol or Flds	(4) Field mas	(5) Resoln mas	(6) Pixels	(7) Sample time sec	(8) Vel Res km/s	(9) Vel Range km/s	(10) Chnls	(11) No. beams	(12) Hyb loops	(13) Deconvo -lution %	(14) Input words	(15) No. of 2-D FFTs	(16) No. of 12 hr /yr	(17) Avg. equiv 2048
H2O Multiple restricted X, Y, V Cubicles																
H2O HII regions	2	100	30.	0.3	300	1.6	0.5	12.	31	0	0	100	0.23E+11	87773	10.	19.
H2O Proper motions	3	40	30.	0.3	300	1.6	0.5	12.	31	0	0	100	0.90E+10	51866	20.	23.
OH																
OH HII multi-fld	2	60	300.	3.0	300	15.0	0.2	3.	19	0	0	100	0.89E+09	32060	20.	14.
OH Supergiants	2	4	3000.	10.0	900	30.0	0.2	60.	390	0	0	100	0.61E+09	33462	4.	32.
OH Miras	2	4	1000.	10.0	300	30.0	0.2	40.	260	0	0	100	0.40E+09	22307	20.	10.
SiO																
SiO Stars	2	4	500.	5.0	300	30.0	0.5	60.	156	0	0	100	0.16E+09	13384	10.	3.
H absorbtion																
Galactic H	3	2	100.	3.0	100	30.0	0.5	200.	520	0	0	100	0.40E+09	22308	10.	0.
Extragalactic	3	2	100.	3.0	100	30.0	0.5	400.	1040	0	0	100	0.99E+09	44616	20.	2.

Percent of year used: 15.7
Total effective FFTs: 103.0

4. - COMPUTING FOR THE PROPOSED MILLIMETER ARRAY

The millimeter array is a future project of NRAO, planned for some time after the VLBA is completed. Its purpose will be to provide a synthesis capability for the wavelength range of 1 to 10 millimeters similar to that now provided by the VLA between 1 and 20cm. Because of differences in the nature of the astronomical sources at millimeter wavelengths, the instrument has somewhat different computer processing needs than the VLA.

First, the ratio of the sizes of the individual elements to the maximum baseline will be larger in the millimeter array than in the VLA. This results in smaller numerical fields of view. The largest numerical fields of view for the Millimeter Array should be a factor of 10 smaller in each coordinate than the largest fields necessary for the current VLA. The sampling interval for observations should also be corresponding longer; however this may depend on exactly how the instrument is used and may be similar to the VLA.

The number of baselines sampled at any one time will likely be less than or equal to the current VLA. However, the bandwidth will be 20 times larger and will be broken into many channels in the correlator even for continuum. Thus the amount of data coming out of the correlator will be similar to the case of the VLA with an extended correlator as discussed in the VLA section. The spectral line data rates should be similar to the VLA. Thus the data rates will be less than and maybe much less than the cases discussed under the VLA.

The Millimeter Array will spend much more of its time near the limits of atmospheric stability and thus will probably need more sophisticated atmospheric corrections and self-calibration. This may impact the computer load but it is difficult to judge the magnitude of the problem at present.

In many cases, the Millimeter Array may make maps of many small fields and then need to mosaic them together. This will require new software but probably will not be a major computing load.

The millimeter region of the spectrum is very rich in spectral lines and thus the Millimeter Array will spend a larger fraction of its time in the spectral line mode and probably more of its time in modes using most of the available channels. In these cases the array will usually not be in its largest configuration so the maps will be even smaller than in the case discussed above. However the maps may tend to be more complex than for the VLA case. Thus the need for user interaction, and computer graphics in smaller machines may be larger than for the VLA.

In summary the computer needs for large fast computers should be small compared with the VLA. Data storage needs could be as large as the

VLA but are probably somewhat less. New calibration techniques could change this but at present the Millimeter Array does not appear to drive NRAO's computing needs.

5. - COMPUTING FOR THE PROPOSED 75MHz ARRAY

It has been proposed that NRAO build a low frequency extension to the VLA, which would operate at a frequency of 75MHz (VLA Scientific Memorandum No. 146, "A Proposal for a Large, Low Frequency Array Located at the VLA Site"). Such an array would have sensitivity two orders better than previously achieved at this frequency and would also provide a substantial increase in resolution to 20 arcseconds. These improvements would allow investigation of a large range of scientific questions for a relatively small cost, only $1,000,000 1984 dollars. The relative inexpense of this array arises from three factors, first, the antennae would consist of banks of simple Yagis, secondly, the VLA waveguide system would be used to transmit signals and from the antennae, and thirdly, since only low bandwidth (2 MHz) continuum observations are envisaged, at least initially, the correlator required is very simple and cheap to build.

The computational load imposed by this array is dependent upon poorly known properties of the ionosphere, which affects the collimation of the wavefronts as they reach the array. It is anticipated that simple extensions of the currently used selfcalibration techniques will suffice to recollimate the wavefronts. The main complication over ordinary selfcalibration is that that, within the power pattern of an antenna, there will a number of regions or "isoplanatic patches" of differing collimation error. Consequently, one complete selfcalibration cycle will required for each patch and the patches will have to be combined into a complete image a number of times. With our current knowledge will estimate that typically there will be about 250 patches, each of one degree diameter, in the antenna power pattern, which will be of diameter 15 degrees. The collimation error due to each patch will vary over a number of minutes, a timescale comparable to that appropriate to the A or B configurations at the higher frequencies. The image size for each patch will be about 512x512 and only full tracks of about 12 hours will be made. Thus, the equivalent load is about 250 512x512 selfcalibrated images, plus some overhead connected with the interaction of different patches, in 12 hours. This load is reduced in two ways : firstly, the 75 MHz array will only operate when the VLA is in A configuration, about 3-4 months per year, and secondly, data taken when the collimation is particularly poor will simply be thrown away. Thus the net, time-averaged load is probably no more than about 100 512x512 selfcalibrated images per 12 hours, which is equivalent to about 8 2048x2048 images per twelve hours. However, this figure is very dependent upon the properties of the ionosphere. Tests to check the typical time- and size- scales are now proceeding.

APPENDIX A

The following discussions have been extracted from VLA Computer Memorandum No. 168, "A Computer Plan for the VLA", by R.Duquet, G.Hunt and R.Burns, September, 1983.

A Canonical VLA Imaging Task

Consider the observing program that entails:

1- An average input data rate of 9,000 complex visibility values per second maintained continuously over 12 hours

2- Simultaneous construction of 256 maps, each defined on a grid of 2048 by 2048 points

3- An output rate capable of storing those 256 maps.

The origin and significance of these specifications can be appreciated by noting that spectral line observations using 256 channels with all of the VLA's 351 baselines generate 89,856 complex visibilities i.e., with the usual 10 second integration period, roughly 9,000 values per second.

A Hypothetical Computer

Consider that the task described above is to be carried out on some machine which is unspecified except for:

1- Its architecture, which is conventional (von Neumann) i.e. it is not an Array Processor.

2- Its processing speed, which is sufficient to perform a 2-dimensional 2048 x 2048 complex FFT in 1 minute. (Roughly equivalent to a processor rating of 20 Mips or, even more roughly, 5 Mflops.)

3- The size of its main memory, a minimum of 8 million words of 4 bytes each (i.e. 32 Mbytes), which is sufficient to hold one copy of an entire 2048 by 2048 complex map (or a map and beam) plus program and ancilliary data.

4- Its I/O (channel) capacity, which can support an effective continuous transfer rate of 1 Mbyte per second (8 Mbaud).

Mapping (Image Formation) on the Hypothetical Machine

The canonical task described above will be considered in two parts: mapping (image formation) and CLEANing (image enhancement).

For a rough estimate of the computational load represented by the mapping part of the reference task, we can make several simplifying assumptions. The most important of these is the assumption that the maps can be produced automatically i.e. unencumbered by the (typically interactive) process whereby the raw data is normally edited. Another assumption is that for mapping the following operations suffice:

1- Application of gain and passband corrections once to each input datum.

2- Gridding by convolution of each datum onto a cell of 6 by 6 grid points.

3- Calculation of a beam pattern for each map.

4- Production of maps. Each map will require a 2-dimensional fast Fourier transform (FFT) of the gridded data.

5- Calculation of another 2-dimensional FFT for each beam. Because the map is Hermitian, this process will not require additional resources since the calculation of the beam can be combined with the calculation of the map.

To apply gain and passband corrections to each datum requires the equivalent of roughly 100 instruction cycles under optimum conditions, i.e. the gain tables have been prepared ahead of time, they are available in main memory and the input data is available in time sequence order. (The equivalence makes allowance for those operations that will be performed on floating-point values.) If we use the approximate figure of 20 Mips for the processor speed then, over the 12-hour reference period, this task will occupy the CPU for

$$\frac{9{,}000 \text{ values/sec} \times 43{,}200 \text{ seconds} \times 100 \text{ instructions}}{20{,}000{,}000 \text{ instructions/sec}} = 0.5 \text{ hours.}$$

Convolving the calibrated data onto 36 grid points will take approximately 750 instruction cycles per datum if the inner loop of this process has been tightly coded. The hypothetical machine will be occupied by the gridding task

$$\frac{9{,}000 \text{ values/sec} \times 43{,}200 \text{ seconds} \times 750 \text{ instructions}}{20{,}000{,}000 \text{ instructions/sec}} = 4.1 \text{ hours}$$

A weight must be calculated for each map point in order to normalize the map and, coincidentally, to determine the beam pattern. In the usual case (uniform weighting) this will require a preliminary pass through the data involving an additional 250 instructions per datum. The hypothetical machine will be occupied by the beam-making task

$$\frac{9{,}000 \text{ values/sec} \times 43{,}200 \text{ seconds} \times 250 \text{ instructions}}{20{,}000{,}000 \text{ instructions/sec}} = 1.4 \text{ hours}$$

The 2-dimensional FFT will be effected by a set of 2048 1-dimensional FFTs in each of two orthogonal directions. (The total number will be 25% less because the Hermitian properties of the data make half of the FFTs in the first direction unnecessary.) Between each set of transforms the grid must be transposed but, since it is assumed that the entire grid can be in memory simultaneously, no allowance has been made for this process.

From the specification that each (full) 2-dimensional FFT would require 1 minute of processing time we find that each 1-dimensional complex 2048-point FFT should take a bit more than 15 milliseconds. So mapping will require:

15 milliseconds/FFT x (1024 + 2048) FFTs x 256 maps = 3.3 hours.

Turning next to the I/O requirements of the task, we start by determining the total number of bytes involved in the input data set and in the output maps. Each input datum is a complex pair whose real and imaginary parts require 2 bytes each. The 12-hour volume of data is therefore:

9000 values/sec x 4 bytes/value x 43,200 secs = 1.5 Gbytes.

In addition to the raw data, a certain amount of overhead is necessary to identify data clusters and to access the data effectively. The identifying information consists of values of u, v, w, t, baseline and flags. Each identifier requires 16 bytes. If a cluster of data (a record) consists of one identifier for each set of 256 channels the overhead is only 16/1024 or less than 1.5%. This is the minimum overhead and the most efficient use of storage.

Unfortunately, the mapping process cannot use data for all 256 channels at one time so, under the identification scheme described above, each cluster would have to be reread many times and the I/O volume would be multiplied by a large factor. A similar problem is encountered if one goes to the opposite extreme of identifying the data for each channel separately. In that case the data storage overhead would be a prohibitive 16/4 or 400%. A reasonable compromise between these two extremes appears to be one identifier per cluster of 8 channels for an overhead of 16/32 or 50%.

The total input data volume is then 2.2 Gbytes.

but for mapping, each value must be read 8 times (once for each channel in the group since we assume only one channel will be mapped at a time).

The total input data volume for mapping is then 17.6 Gbytes.

The output data are the pixels in the 256 maps. If each pixel is stored in a 4-byte floating point number the output data volume is:

256 maps x (2048 x 2048) pixels x 4 bytes/pixel = 4.3 Gbytes.

Properly identifying maps does not usually entail nearly as much overhead as that required for the identification of the input data. In fact such overhead for maps can be neglected. What we should not neglect, however, is the storage requirement for beams. Since the canonical task deals with spectral line data, no more than 10% of the beams will be needed. Furthermore, since beam patterns are symmetrical, each one will require only half as much space as the corresponding map. So the need to store beams raises the 12-hour volume of output to

4.3 Gbytes + 0.2 Gbytes = 4.5 Gbytes.

Since our hypothetical machine can perform continuous high speed I/O, averaging a full Mbyte per second (8 Mbaud) over the entire 12-hour period, the time required to pass all the input data through the I/O channels will be 2200 seconds (0.6 hours), the input for mapping will be 8 times that much (4.8 hours), and the time required to write the 256 maps and beams will be 1.2 hours.

It is unlikely, however, that the input data can be used synchronously i.e. that it can be processed as soon as it is available and that it can be discarded thereafter. The need to buffer the input data means that it must pass through an I/O channel at least 3 times: once to acquire it from the original source, again to store it on a disk and a third time to read it back when it is needed in the map-making process. As a matter of fact, the standard practice for gridding (use of uniform weighting) requires that the data be read twice from the disk.

The output must be transfered asynchronously to a display system which means that it too must be stored (temporarily at least) on a disk.

In summary, here is how a 5 Mflops machine, with 1 Mbyte/sec average I/O bandwidth and somewhat more than 32 Mbytes of memory available for the process, would handle the canonical imaging task:

	Operation	Hours
I/O	Original Input	0.6
	Output to Buffer Disk	0.6
	First Input to Mapping	0.6
	Second Input to Mapping	4.8
	Original Map and Beam Output	1.2
	Rereading Output from Disk	1.2
	Writing Maps & Beams to Tape	1.2

		10.2
CPU	Calibration	0.5
	Weight Determination	1.4
	Gridding	4.1
	FFT	3.4

		9.4

Thus, even with fully overlapped I/O and computation, with no CLEANing (let alone self-calibration), with no display, no calculation of gain tables, and no editing, the mapping part of the canonical task would come close to saturating the hypothetical 5 Mflops machine, especially if the operating system overhead were a typical 15 to 20% of application program time.

The figures quoted above are obviously only rough estimates. It may not be quite so obvious that each approximation has been biased in the direction of underestimating rather than exaggerating the computer load. For example, in addition to all of the exclusions listed in the previous paragraph, the requirements have been underestimated further by the following:

1. Data is assumed to be available in main memory whenever it is needed. Inadequate memory would certainly increase the time estimates by a significant factor. In particular, the time to transpose the data between the two sets of FFTs would no longer be negligible.

2. The CPU time required by programs that transfer data from the original source to the disk buffer (the FILLER program), to copy the output maps from disk and reformat them for tape (the FITS program), and to write a calibrated version of the database onto tape for export to other installations (the UVFITS program) have all been neglected.

3. It has been assumed that the data could be organized to allow streaming at 1.0 Mbyte per second into and out of the central processor. In particular, it has been assumed that no sorting is required.

4. A multi-user environment would (presumably) require paging and greatly increased system overhead. An interactive environment (even with a single user) would also slow down the process by the amount of time the system would be waiting for user responses.

5. No provision has been made for archiving either the original data or the maps (i.e. neither BACKUP nor RESTORE have been included in the I/O requirements).

The only way in which the necessary resources may have been overestimated is by the assumption that all of the input data will be used to make maps. This ignores the phase, amplitude and passband calibrators which, in general, will not be mapped. However these sources will be used for their respective purposes and will have their own impact on system resources.

Increasing Dynamic Range (Image Enhancement)

The second part of the reference task is CLEANing (image enhancement). This computer-intensive deconvolution procedure, together with the relatively new technique of self-calibration, has upgraded the VLA from an instrument with 100:1 dynamic range into one that routinely achieves 1,000:1 and in some cases even 10,000:1 dynamic range. In other words, the value of VLA data has been increased many times over at the expense of a multifold (and largely unanticipated) increase in the amount of computer power required.

Tim Cornwell has calculated the amount of computer processing required by the CLEAN algorithm in terms of an equivalent number of 2-dimensional FFTs. In a memo distributed informally within the VLA (reproduced here as Appendix D) Cornwell estimates that

1 CLEAN deconvolution = (2 + 2 N) 2-dimensional FFTs

where N is the number of "major cycles" required to achieve the desired dynamic range. Typical values of N range from 4 to 200 but fortunately, for spectral line work, the preponderant value is 4. Even so, this means that it takes at least 10 times as long to CLEAN a map as to make the map initially. On the hypothetical machine considered above, CLEANing the 256 maps (12 hours of data) would take

256 maps x 10 FFTs/map x 1 min/FFT = 2560 minutes = 43 hours !!

It is because CLEANing is so dominant that it has been considered separately from the rest of the canonical task described above!

The CLEAN algorithm requires that the initial map and beam be reread from the disk. Obviously, we will want to save the improved map so it must be written to disk. In summary, the need to CLEAN the 256 maps (2048 by 2048) produced by the canonical task described above imposes the following load on the hypothetical machine:

	Operation	Hours
I/O	Read Maps and Beams from Disk	1.8
	Output CLEANed Maps	1.2

		3.0
CPU	10 FFTs per channel	43.0

Self-calibration (automatic recalculation of antenna gains) is another computer-intensive process that is used in the quest for ever higher dynamic range.

Tim Cornwell has estimated the computer load represented by a self-calibration step. It turns out to be approximately the same as one 2-dimensional FFT followed by a complete re-mapping. The actual redetermination of gain values is relatively insignificant.

Although self-calibration is a large percentage of the work required by continuum observations, it is a small factor in spectral line work because the gain figures obtained by self-calibrating one channel can be used to map all of the other 255 channels. So, for the canonical task, self-calibration can be ignored.

Summarizing the total computer requirements of the canonical task:

	Operation	Hours
I/O	Mapping	10.2
	CLEANing	3.0

	TOTAL	13.2
CPU	Mapping	9.4
	CLEANing	43.0

	TOTAL	52.4

The conclusion that emerges from these calculations is that the computer requirements of the canonical task exceed the capacity of the hypothetical machine by a factor of 4 or 5. In other words, a machine of 20 to 30 Mflops would be needed merely for image formation to process the data at the rate it is observed. Real circumstances, in which the ideal assumptions made in these calculations were not all valid, plus ancilliary tasks (such as data editing) would raise the level of computer power required to handle the canonical task as an average load to about double that figure or to about 40 or 50 Mflops.

It now remains to be shown whether or not the canonical task is representative of the anticipated work to be done at the VLA during the next decade.

APPENDIX B

The work of Ekers et al. is absolutely fundamental to this analysis of VLA computer requirements. Given a fixed set of objectives and the computational effort required to meet each, the size of the machine needed by the VLA follows. (Naturally, professional judgment also influences the size calculations.) Different objectives, of course, would lead to different computer requirements.

The processor load for each research area was first obtained by using Cornwell's equations to determine the equivalent number of 2-dimensional Fourier transforms. The number of data points in this transform was taken to be the appropriate number of pixels on each side of the map as given by Ekers et al. In any actual experiment the map size would probably be a power of 2 but the numbers derived by Ekers et al. were used as given, on the grounds that some users would ask for the next larger size while others would make do with the next smaller size. The number of equivalent 2-dimensional 2048 by 2048 complex FFTs (interpretable as minutes of processing time on the hypothetical 5 Mflops machine) was obtained by scaling according to

$$N^2 \, LOG(N).$$

Ekers et al. have estimated the percentage of all VLA observations that would be made for each of the experiment types. Those estimates were used to combine the calculated requirements for each experiment type in a weighted average that represents the computer requirements to process a "typical" 12 hours of VLA observing.

For continuum observations the amount of work required by each of the experimental areas was estimated from the following equations:

$$Nfft = Obs * (DC + 30 * SC\%)$$

where

Nfft is the number of 2D FFTs (or equivalent) to be performed. The number of data points in each FFT is that appropriate for the experiment i.e. has not yet been scaled to minutes of processing on the hypothetical 5 Mflops computer.

Obs is the number of sources observed in this type of experiment during a typical 12-hour period.

DC is the work required for mapping and for CLEANing each source. It is obtained from

$$[Np/2 + (Np-1) * 12 * C\%] * R$$

Np is the number of polarizations observed. The constant 2 reflects the fact that, while some experiments require a separate map for each IF, others require only one. On average, the number of maps should be about half the number of IFs observed.

C% is the percentage of maps to be CLEANed. (The constant 12 is based on the belief that, on average, only 5 major cycles are required to CLEAN each CONTINUUM map to the desired level.)

R is a repetition or "re-try" factor. It is an estimate of the number of times a given data set might be processed.

SC% is the percentage of maps to be self-calibrated. The constant 30 assumes that self-calibration will be carried out for an average of 2.5 iterations. Approximately, the iterations cost one FFT for the initial map, 2+2*N FFTs for CLEAN, followed by 2 FFTs for the self-calibration step itself. N, the number of major CLEAN cycles, is expected to be approximately 2, 4, 6, etc. for successive iterations. The total equivalent FFT load for the average self-calibration task, therefore, is

$$(1+6+2) + (1+10+2) + (1+14+2)*0.5 = \text{(app) } 30$$

Table 3A (on the following page) shows the result of applying these equations to the values supplied by Ekers et al. for CONTINUUM experiments. The leftmost columns are labelled with symbols that have been used in the equations. The labels on the righthand columns carry the following meaning:

Time = (Equiv + 3.3 * Vis) * Mix

where

Time — is the number of minutes of processing time required on the hypothetical 5 Mflops machine for the given type of experiment in a "typical" 12-hour mix of CONTINUUM experiments.

Equiv — is the equivalent number of 2048 x 2048 2-dimensional complex FFTs. It is given by

$$\text{Equiv} = \frac{(\text{Npix} * \text{Npix}) * \text{LOG}(\text{Npix})}{(2048 * 2048) * \text{LOG}(2048)} * \text{Nfft}$$

Npix — is the number of pixels on each side of the maps required for the given experiment.

Vis — is the number of words of visibility data read in this experiment (in millions of words per 12 hours). The constant 3.3 is the number of minutes required to grid a million words of data.

Mix — is the estimated percentage of CONTINUUM observing at the VLA during which the given type of experiment is conducted.

Table 3A - Processing time on a Hypothetical 5 Mflops Machine for the anticipated 12-hour mix of CONTINUUM observations

Type of Experiment		Obs	Np	R	C%	SC%	Npix	Nfft	Equiv 2K	Vis Mw	Mix %	Time min
Full field mapping												
A array												
	1.4 GHz	6	4	2	100	75	5400	591	4762	400	4	190
	5 GHz	6	4	2	100	75	6000	591	5918	400	5	295
	15,22 GHz	6	4	2	100	75	5400	591	4696	200	2	93
B array												
	1.4 GHz	6	4	2	100	75	1600	591	391	130	5	19
	5 GHz	6	4	2	100	75	1800	591	471	70	3	14
	15,22 GHz	6	4	2	100	75	1600	591	354	16	1	3
C,D arrays												
	1.4 GHz	6	4	2	100	75	500	591	39	32	2	0
	5 GHz	6	4	2	100	75	600	591	45	8	5	2
	15,22 GHz	6	4	2	100	75	500	591	30	4	3	0
Small Objects												
Weak Sources	327 MHz	12	2	2	100	100	9400	672	16993	23	6	1019
	1.4 GHz	6	4	2	100	50	5400	546	4287	26	6	257
	5 GHz	6	4	2	100	50	3000	546	1232	7	5	61
	15,22 GHz	6	4	2	100	50	300	546	10	4	4	0
Strong Sources (all)		6	4	2	100	100	300	636	11	4	5	0
Snapshots												
Weak Sources	1.4 GHz	100	3	1	100	50	5400	4050	31741	13	4	1269
	5 GHz	100	3	1	100	50	3000	4050	9126	2	5	456
	12,22 GHz	100	3	1	100	50	300	4050	65	2	4	2
Strong Sources (all)		200	4	1	100	100	300	13600	219	4	9	19
Survey												
B array,	1.4 GHz	720	2	1	100	20	1100	13680	3625	3	1	36
C array,	1.4 GHz	4320	2	1	20	0	500	14688	714	2	2	14
Point Sources												
Astrometry (A array)		100	2	1	0	0	100	100	0	2	3	0
Monitoring, Spectra		100	4	1	0	50	100	1700	3	2	2	0
Flare Stars		10	4	1	0	0	100	20	13	40	1	0
Detections		24	2	1	0	0	100	24	0	2	7	0
Solar												
Quiet		2	4	1	100	0	500	76	5	4	1	0
Active		20	4	1	100	0	500	760	63	81	3	1

Time to process 12 hours of "typical" CONTINUUM observations: 63 hours

Visbilities for "typical" CONTINUUM observations: 58 Mwords / 12 hours

For spectral line observations the amount of work required by the experimental mix foreseen for the VLA was estimated from the following equations:

Nfft = Obs * [(Np-1) * (Ch * 1.3) * (1 + B% + 10 * DC%) + 15 * S]

where the symbols are mostly as defined for the CONTINUUM case, i.e.

Nfft — is the number of 2D FFTs (or equivalent) to be performed. The number of data points in each FFT is that appropriate for the experiment.

Obs — is the number of sources observed in this type of experiment during a typical 12-hour period.

Np — is the number of polarizations observed.

Ch — is the number of channels observed. The constant 1.3 reflects the belief that, on the average 10% of the channels will be mapped 3 times before the full set is mapped.

DC% — is the percentage of maps to be deconvolved (CLEANed). The constant 10 is based on the belief that, on average, only 4 major cycles are required to CLEAN a SPECTRAL LINE map to the desired level.

B% — is the percentage of channels for which beams are mapped.

S — is a constant which has the value 0 or 1. It will be 1 if a map (for 1 channel at most) is to be self-calibrated, otherwise it will be 0. The constant 15 is an estimate of the work required for self-calibration when it is required.

Table 3B (on the following page) shows the result of applying these equations to the values supplied by Ekers et al. for SPECTRAL LINE experiments. The leftmost columns are labelled with symbols that have been used in the equations. The labels on the righthand columns carry the following meaning:

$$\text{Time} = (\text{Equiv} + 3.3 * \text{Vis}) * \text{Mix}$$

where

Time is the number of minutes of processing time required by this type of experiment in a "typical" 12-hour mix of all types of CONTINUUM experiments.

Equiv is the equivalent number of 2048 x 2048 2-dimensional complex FFTs. It is given by

$$\text{Equiv} = \frac{(\text{Npix} * \text{Npix}) * \text{LOG}(\text{Npix})}{(2048 * 2048) * \text{LOG}(2048)} * \text{Nfft}$$

Npix is the number of pixels on each side of the maps required for the given experiment.

Vis is the number of words of visibility data read in this experiment (in millions of words per 12 hours). The constant 3.3 is the number of minutes required to grid a million words of data.

Mix is the percentage of all SPECTRAL LINE observing at the VLA during which the given type of experiment is conducted.

Table 3B - Processing time on a Hypothetical 5 Mflops Machine for the anticipated 12-hour mix of SPECTRAL LINE observations

Type of Experiment	Obs	Np	Ch	DC%	S	B	Npix	Nfft	Equiv 2K	Vis Mw	Mix %	Time min
Extragalactic Emission												
HI Clusters	1	2	16	20	0	12	1100	64	24	21	6	1
HI Galaxies	2	2	64	75	0	3	500	1419	90	65	12	10
ISM (individ. gal)	3	2	130	50	0	15	5400	3118	24712	850	6	1482
Recomb. line	1	2	8	50	1	1	200	77	3	8	1	0
NH3	1	2	8	10	0	0	100	20	2	8	1	0
H2O Masers	3	4	4	0	0	7	6000	50	509	58	1	5
OH Masers	3	4	64	0	0	6	5400	793	6495	839	1	64
Extragalactic absorption												
HI	3	2	64	50	1	2	400	1547	67	65	2	1
OH	2	4	65	50	1	2	400	3082	135	130	3	4
H2CO	2	2	65	50	1	7	1200	1055	368	95	2	7
Galactic Center												
NH3	2	2	16	50	1	0	100	279	5	16	3	0
H2CO	1	2	64	50	1	3	5400	516	4186	419	1	41
HI, OH	1	2	256	50	1	3	5400	2021	16400	1700	1	164
Recomb. Line	1	2	8	50	1	4	1600	77	51	16	1	0
Galactic Absorption												
HI	3	4	130	100	1	3	5400	16821	132377	1700	2	2647
OH	3	4	130	100	1	1	1800	16791	12938	570	2	258
H2CO	3	2	130	100	1	3	6000	5637	55517	950	2	1110
NH3	1	2	128	100	1	0	6000	1845	18377	930	4	735
Galactic Emission												
HI	2	2	52	50	0	1	400	812	41	52	2	0
NH3	1	2	128	10	1	0	100	347	43	130	12	5
Recomb Lines	2	2	65	50	1	0	100	1044	23	66	11	2
OH Masers	12	4	130	50	1	1	1800	36744	28090	570	7	1966
H2O Masers	2	4	32	5	1	2	6000	409	4163	469	6	249
Stars												
OH Masers	12	4	156	50	1	0	100	43984	168	320	6	10
H2O Masers	12	4	32	50	1	0	600	9165	681	65	3	20
SiO Masers	12	4	16	50	1	0	1200	4672	1507	47	2	30

Time to process 12 hours of "typical" SPECTRAL LINE observations: 147 hours

Visbilities for "typical" SPECTRAL LINE observations: 314 Mwords / 12 hours

Table 3C - Processing time on a Hypothetical 5 Mflops Machine for the anticipated 12-hour mix of SPECTRAL LINE observations ASSUMING AN ENHANCED CORRELATOR

Type of Experiment	Obs	Np	Ch	DC%	S	B	Npix	Nfft	Equiv 2K	Vis Mw	Mix %	Time min
Extragalactic Emission												
HI Clusters	1	2	325	20	0	12	1100	1318	490	430	6	29
HI Galaxies	2	2	130	75	0	3	500	2883	182	130	12	21
ISM (individ. gal)	3	2	130	50	0	15	5400	3118	24712	850	6	1482
Recomb. line	1	2	130	50	1	1	200	1030	49	130	1	0
NH3	1	2	325	10	0	0	100	844	109	330	1	1
H2O Masers	3	4	325	0	0	7	6000	4068	41386	4700	1	413
OH Masers	3	4	325	0	0	6	5400	4030	32995	4300	1	329
Extragalactic absorption												
HI	3	2	130	50	1	2	400	3097	135	130	2	2
OH	2	4	65	50	1	2	400	3082	135	130	3	4
H2CO	2	2	65	50	1	7	1200	1055	368	95	2	7
Galactic Center												
NH3	2	2	520	50	1	0	100	8142	185	530	3	5
H2CO	1	2	520	50	1	3	5400	4091	33175	3400	1	331
HI, OH	1	2	520	50	1	3	5400	4091	33175	3400	1	331
Recomb. Line	1	2	325	50	1	4	1600	2566	1725	640	1	17
Galactic Absorption												
HI	3	4	130	100	1	3	5400	16821	132377	1700	2	2647
OH	3	4	130	100	1	1	1800	16791	12938	570	2	258
H2CO	3	2	130	100	1	3	6000	5637	55517	950	2	1110
NH3	1	2	416	100	1	0	6000	5963	59388	3000	4	2375
Galactic Emission												
HI	2	2	52	50	0	1	400	812	41	52	2	0
NH3	1	2	416	10	1	0	100	1096	139	419	12	16
Recomb Lines	2	2	65	50	1	0	100	1044	23	66	11	2
OH Masers	12	4	130	50	1	1	1800	36744	28090	570	7	1966
H2O Masers	2	4	260	5	1	2	6000	3112	31727	3800	6	1903
Stars												
OH Masers	12	4	156	50	1	0	100	43984	168	320	6	10
H2O Masers	12	4	156	50	1	0	600	43984	3272	320	3	98
SiO Masers	12	4	156	50	1	0	1200	43984	14193	460	2	283

Time to process 12 hours of "typical" SPECTRAL LINE observations: 228 hours

Visbilities for "typical" SPECTRAL LINE observations: 835 Mwords / 12 hours

The estimates of future research levels at the VLA are based upon a great deal of painstaking work and many hours of discussion with numerous well informed observers. Nevertheless, the quantitative results in Table 3 must be considered to be no better than rough estimates. The following are some of the sources of uncertainty in these values:

1- They are neither "worst case" nor "average" but somewhere in between; call them "mildly severe" cases. It is not possible to quantify that description very closely.

2- They assume a perfectly distributed work load (i.e. slack time consistently filled by backlogged work). The wasteful effects of uneven use could easily double the computer power needed for busy (or even normal) periods.

3- The introduction of new mathematical procedures that are more demanding computationally than those presently in use could easily double or even quadruple future VLA computer needs.

Such uncertainties must be kept in mind as we examine the implications of Table 3.

If we assume that there will be roughly equal amounts of continuum and spectral line observing at the VLA, the time needed to process 12 hours of observations on a 5 Mflops machine would be 109 hours! Looking at this result in a different way, the mix described by Ekers et al. calls for a machine of at least 45 Mflops (60 Mflops if an enhanced correlator is assumed) **merely for image formation and enhancement.** A practical machine to meet these and other concomitant needs in a non-ideal environment would have to be capable of delivering more computer cycles, i.e. it should be rated at no less than 60 to 75 Mflops.

....

Some rather important characteristics of the estimated VLA computer requirement can be observed by ranking all anticipated observing programs according to the contribution each makes to the overall total. In Table 4 the cumulative sum of the individual requirements is shown along with the cumulative percentage of observing time expected for these programs.

Table 4A
Current Correlator

% obs	Mflops	% Mflops
100.0	43.7	100.00
99.0	34.5	78.96
95.5	27.7	63.33
92.5	22.5	51.55
90.5	18.1	41.46
89.5	14.3	32.64
86.5	10.7	24.54
84.5	8.2	18.69
82.0	6.6	15.07
79.5	5.6	12.72
78.5	4.7	10.66
75.5	3.8	8.62
72.5	2.9	6.63
70.5	2.2	5.12
70.0	1.7	3.81
69.0	1.3	3.07
68.5	1.1	2.55
66.0	0.9	2.06
65.5	0.8	1.73
65.0	0.6	1.44
64.0	0.5	1.20
62.5	0.5	1.04
58.0	0.4	0.88
55.5	0.3	0.73
54.5	0.3	0.61
53.0	0.2	0.50
47.0	0.2	0.41
44.0	0.1	0.33
43.0	0.1	0.27
37.0	0.1	0.23
36.5	0.1	0.19
35.0	0.1	0.16

Table 4B
Enhanced Correlator

% obs	Mflops	% Mflops
100.0	60.5	100.00
99.0	51.3	84.80
97.0	43.0	71.16
93.5	36.2	59.87
90.5	29.6	48.94
87.5	24.4	40.43
85.5	20.0	33.14
84.5	16.2	26.76
81.5	12.6	20.91
79.0	11.1	18.29
78.5	9.6	15.91
78.0	8.5	14.00
77.5	7.3	12.10
77.0	6.2	10.21
74.5	5.1	8.51
73.5	4.2	6.88
72.5	3.3	5.39
69.5	2.4	3.91
67.5	1.7	2.82
66.0	1.4	2.26
65.0	1.0	1.72
62.5	0.8	1.36
62.0	0.7	1.16
59.0	0.6	0.99
53.0	0.5	0.86
48.5	0.5	0.75
46.0	0.4	0.63
45.5	0.3	0.54
39.5	0.3	0.44
38.5	0.2	0.36
37.0	0.2	0.28
34.0	0.1	0.22

Twenty additional experiment types require, in aggregate, 0.1 Mflops or less.