National Radio Astronomy Observatory
Charlottesville, Virginia

14 July 1984

To: VLBA Computer Coordination Committee

From: R. C. Walker

Subject: Need for a Large Computer

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 Ekers, Fomalont, and Owen (1983, VLA Scientific Memorandum No. 150 - EFO). They conclude 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, they present tables 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 VLBA.

This memo is an attempt to specify the computing needs of the VLBA in a manner similar to that used by EFO. 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 VLBA needs help justify the aquisition of the large computing facility. The breakdown of the projected computing needs by class of observations is given in Tables I and II 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 EFO 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 Duquet, Hunt, and Burns (VLA Computer Memorandum No. 168), a 1 Mflop computer could calculate about 144 2048 by 2048 FFT's per 12 hours (the units of compute power used in EFO 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 EFO. 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 EFO 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 EFO. 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 EFO 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). EFO 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 EFO 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)

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 EFO. 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 EFO.
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). EFO 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 EFO 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)

(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

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
H20 Multiple restricted X, Y, V Cubicles																
H20 HII regions	2	100	30.	0.3	300	1.6	0.5	12.	31	0	0	100	0.23E+11	87773	10.	19.
H20 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