VLBI ARRAY COMPUTER USEAGE

3 JULY 80

W. Cotton and J. Benson

VLB ARRAY MEMO No. 4

I. Introduction

Both computing hardware and data analysis are developing sufficiently rapidly that accurate predictions for the computing needs of the VLB array are not possible. This report will, instead, examine the requirements of a ten antenna array on computing facilities currently available in Charlottesville.

II. Post-correlation Processing.

A. Continuum

1) In the current Charlottesville system the first post-correlation step is to correct the data for peculiarities in the processor and to pre-average the 0.2 second integrations from the correlator to several seconds in order to reduce the volume of the data. In newer correlators such as those at Cal-Tech and Haystack much of this step is done on-line by the correlator/computer. In Charlottesville this is done in the IBM 360.

2) Following pre-averaging is the fringe fitting step which results in the estimates of the complex correlation coefficient, group delay and fringe delay rate from data coherently averaged for several minutes. This step is also done in the IBM 360 in Charlottesville.

3) After fringe fitting the data is edited, removing bad data.

4) Current practice is to calibrate the correlation amplitudes to Janskys using measured system temperatures and antenna sensitivities. Except in certain phase referencing experiments phase calibration is considered hopeless at this stage and is ignored.

5) Finally the phases are iteratively calibrated and the map produced by a technique known as self calibration or hybrid mapping. In this technique the current model of the source (CLEAN point components or the initial guess) is used to calibrate the phases relative to an arbitrary position in the sky. This technique makes use of the fact that the number of calibration phases needed is N-1 and the number of observed phases is N*(N-1)/2 where N is the number of antennas involved. Similar amplitude calibration is also possible and likely necessary for 1.3 cm observations. For VLBI observations this step is currently done on the Charlottesville VAX using the

Cal-Tech VLBI package but in the future the VLA package will almost certainly be used. Thus the Fred Schwab's self-calibration program was used for timing purposes for this report.

B. Spectral Line

1) Program DECODE preaverages the correlator output.

2) Program AVERAGE corrects clock drifts, residual phase delay rates and makes instrumental phase corrections; then averages both the cross and auto correlations.

3) Program BOG transforms the auto correlations to obtain source spectra and then corrects auto and cross correlations for effects of the bandpass

4) Program CVEL corrects the data for the Earth's motion.

5) Program CAL calibrates the cross correlations using the total power spectra obtained from the autocorrelations.

6) Program PHSREF calibrates phases relative to a reference feature.

7) Program SWAMP, used in editing the data, displays coherent fits to each spectral channel; usually used several times.

8) Programs SWAMP,JANET and DUNE produce a fringe rate map prior to aperature synthesis.

After the above programs are run, the data are fully calibrated and can be analysed by VLA spectral line software.

III. Computing requirments

A. Continuum

Computing requirments for various stages of processing were determined from timing the processing of sample data in the appropriate computer. Correction, preaveraging and fringe fitting were done in the IBM 360; the CPU time requirments for these steps are shown in Table 1.

Table 1
Continuum: pre - Mapmaking (IBM 360 CPU times)

Step	Baseline hour	Array hour (10 ant.)
correction and preaveraging	0.80 min.	36.1 min.
fringe fitting	0.45 min.	20.5 min.
Total	1.25 min.	56.6 min.

Editing and initial amplitude calibration use so little CPU time that their requirments are negligible compared to the uncertainty in the other steps and are thus ignored. However, editing and calibration are the stages which require the most user interaction and are normally repeated several times. If in the future amplitude calibration is done using VLA software initial amplitude calibration may become non-trivial in the computing budget (see the self calibration description for a crude estimate of the time requirments).

To determine the time requirments for self calibration, a VLA data set using 10 antennas with 35,763 data points was used. This is approximately tha amount of data in 12 hours of 1 minute integrations from a ten antenna array. For this test, Fred Schwab's self calibrations program was used to do both amplitude and phase self calibration. The REAL run times for the MODCOMP were recorded and the results are summarized in Table 2. For comparison IBM 360 CPU times for the same mapping and CLEANing tasks are also shown in Table 2.

It is not clear how many iterations through the self calibration procedure will be necessary for 10 antenna data but 10 is probably a safe guess.

TABLE 2
Continuum: Mapping and CLEANing

Step	MODCOMP real time	IBM 360 CPU time
Maping 256X256 cells	5.5 min.	6.2 min.
CLEAN 127X127 500 comp.	8.3 min.	8.2 min.
Self calibration (full complex)	17.2 min.	
Total (1 pass)	31.0 min.	
Per source (10 passes)	310. min. = 5.2 hrs.	

B. Spectral Line.

The CPU time required for programs 1-8 were determined from sample data processed on the IBM 360, These results are summarized in Table 3.

TABLE 3
Spectral Line: pre - Mapmaking (IBM 360 CPU times)

STEP	Baseline hour	Array hour
DECODE	0.80 min	36.0 min
AVERAGE	0.52 min	23.4 min
BOG	0.21 min	9.5 min
CVEL	1.64 min	73.8 min
CAL	0.48 min	21.6 min
PHSREF	0.83 min	37.4 min
SWAMP	>0.7 min	>31.0 min
SWAMP, JANET, DUNE	0.11 min	5.0 min
TOTAL	5.29 min	237.6 min = 4.0 hr

Since VLA spectral line software is not yet avaliable, an estimate of the computing requirments must be made from the continuum test. The mapping and CLEANing could be done as for continuum maps but separately for each frequency channel. In the following estimates 256 spectral channels were assumed; Table 4 shows these estimates.

TABLE 4
Spectral line: MODCOMP real time

STEP	MODCOMP real time
Map making (256 X 256 X 256)	1408. min = 23.5 hr
CLEAN (127 X 127 X 256) 500 comp.	2125. min = 35.4 hr
Total	3533. min = 58.9 hr = 2.5 day

IV. Hardware and Software Development.

A. Continuum

If VLA software is used for editing and calibration then no independent software development past the fringe fitting stage will be necesary. The speed of the correlator correction, pre-averaging and fringe fitting could be improved up to a factor of 10 with the use of an Array Processor. In view of the large amount of CPU time required for these steps, an Array Processor would appear to be vital. In addition, the development of more sophisticated calibration and fringe fitting techniques may be able to take advantage of the large number of baselines to make substantial improvments in sensitivity over the current techniques. However, such techniques are likely to significantly increase the computing requirments.

The MODCOMP self-calibration programs described above make limited use of an array processor. These programs will undoubtedly be improved; for instance, an experimental version of a CLEANing program of the type designed by B. Clark did an equivalent CLEAN to the one described in Table 2 in 1.4 min. MODCOMP real time.

B. Spectral Line

Many of the programs described in Section II could be adapted to use with an array processor, reducing the enormous computing load. Past the calibration stage, VLA software will be completely adequate and no independent development is necessary. Since spectral line mapping and CLEANing for VLA data produce similar problems it is likely that a more efficient method than the one discussed here will be developed.

Appendix to VLBA Memo #4 Computer Useage :

or How much is that in real money?
W. Cotton and J. Benson 28 Aug 1980

The purpose of this appendix is to update VLBA memo #4 and to state the estimated post-correlation computer requirments. As in all such estimates, the numbers quoted here should be viewed with some skepticism.

I. Continuum Fringe Fitting.

The values given in VLBA memo #4 in Table 1 were based on Mk II values and are not entirely relevant to the VLBA. Estimates of Mk III fringe searching are based on values obtained from A. Rogers at Haystack from their experience with the Mk III correlator. One example:

8.5 min of data searched, 2 sec preaveraging, 14 tracks
Disk I/O = 40 sec
mostly FFT search = 120 sec.
search window = 0.8 Hz and 0.5 microsec.

The figures given above were based on a firmware FFT which does a full complex 1024 point FFT in 200 millisec.; the advertised time for the same operation on an FPS array processor is 6 millisec. A crude conversion of the above values to the useage on a machine with an array processor is 0.25 CPU-Hr / baseline hour; or 11 such systems to keep up with the processor. It should be noted that the time is 80% disk I/O. This requirment can be reduced in several ways: 1) the fringe search window can be restricted and data preaveraged for longer times. If the amount of I/O can be reduced by a factor of 10 then one computer plus array processor will be sufficient. 2) In the case of strong sources, either the amount of data recorded can be reduced and/or only a subset of the data used for the fringe search.

II. Useage Requirments.

In order to estimate the amount of computing required Table 1A uses the values given above and in VLBA memo #4. The requirments are expressed in units of a minicomputer (such as a VAX) with an array processor.

Table 1A
Total Computer Useage Requirments

Process	No. Minicomputer + Ap
Cont. pre-mapping	1
Cont. mapping	0.5
Spectral pre-mapping *	1
Spectral mapping *	1
Total	3.5

* Spectral line observations assumed 20% of the time.