VLA COMPUTER MEMORANDUM NO. 140

VLA POST-PROCESSING: PHASE I

W. R. Burns, E. W. Greisen

March 25, 1977

I. Introduction

In this memorandum we discuss in some detail our ideas concerning what we call Phase I in the development of a VLA post-processing capability. Phase I has already begun and will continue, we anticipate, until 1980. This phase is characterized by low levels of personnel and purchasing commitments and by the use of existing facilities in Charlottesville including the IBM 360/65 and the Modcomp IV mini computer. We divide this memorandum into three parts: current accomplishments, short-term goals, and longer range plans. Phase II, the development of post-processing in the 1980's, is not addressed here.

II. Current Accomplishments

The first goal in the Phase I effort is to provide some support for users of the VLA. In the early stages of Phase I this support will be limited to procedures similar to those in use with the Green Bank interferometer. Of course, as the more advanced capabilities discussed in Section IV of this memorandum are developed, they will be made available to users as well.

At the present time we have developed for the IBM 360/65 a package of PL/1 programs which can at a single frequency per pass, read, edit, calibrate, display, and map the output of the Synchronous Subsystem of the VLA. The adopted format allows for data from as many as 351 baselines. User control of the programs is via Job Control Language (for data sets), PARMS cards (for option selection), and INCLUDE/EXCLUDE cards (for data subset selection). The last of these provides a particularly powerful and versatile tool for editing the data throughout the many processing stages. The program VLATOIBM is used to convert the Synchronous Subsystem tapes into our format and may be used for preaveraging and minor editing and concatanation operations. VLAINDEX is used to provide a summary of the data set. VLACLEAN is an editing program which employs user-set levels in automatically deleting data affected by shadowing or identified as marginal by the reliability flags set by the Synchronous Subsystem. It may also be used to delete data points whose amplitudes deviate excessively from the mean. VLACOPY is used to copy data sets, or to carry out "manual" editing via the INCLUDE/EXCLUDE cards. VLAPOS adjusts phases for changes in phase reference position. VLACORR corrects phases for atmospheric effects and for corrections in antenna station coordinates and corrects amplitudes for elevation effects. VLABCAL solves for corrections to the antenna station coordinates and, optionally, to the positions of the calibration sources. VLACAL fits a time varying complex instrumental gain function to data on calibration sources and corrects the data. VLAPCAL determines the instrumental polarization parameters and corrects the data. The observations may be averaged to selected integration times using VLAAVE, and they may be sorted on a wide variety of parameters using VLASORT. A flexible selection of printer displays are provided by VLASCRIB and VLAVIST. These include plots and a choice of compact or expanded printouts of visibility data as a function of time, printouts of visibility data as a function of projected spacing, and plots of one selected parameter versus another with a wide variety of selections available. There are two u-v plane model fitting programs: VLAPTSRC for unconfused point sources and VMODFIT for models consisting of sums of elliptical gaussians and for the polarizations of such models. The program VLAMODEL allows these models to be added to, or to replace, the observed visibility functions in the data set. VLAMAP sorts, grids, and Fourier transforms the data. A number of map processing routines are also provided including VMCLEAN (for the Högbom CLEAN procedure) VMCOPY (for moving and editing map data sets), VMINDEX (for listing map data sets), VLAPRINT (for printer displays), VLADRAW (for Calcomp contour maps), and VLAPIX (for photographic representations).

These programs and formats are documented in the "User's Guide for VLA Data Reduction in Charlottesville", NRAO Users Manual Series Number 29 by Eric W. Greisen and Frederic R. Schwab. Background information concerning the use of data sets and the Job Control Language and extended discussions of some of the algorithms may be found in "User's Guide to 360/65 Interferometer Data Processing Programs", by L. C. Blankenship, R. M. Hjellming, B. L. Meredith, and F. R. Schwab. These two manuals are maintained on Pandora text editor files in Charlottesville and copies are available upon request.

Some rough edges do remain in the programs and their documentation. This is in part due to our present limited understanding of the instrument. At this writing, the principal problems are:

(1) final implementation and testing of the atmospheric and elevation corrections in VLACORR,

(2) detailed testing of VLAPCAL, and

(3) completion of the documentation.

In fact, the User's Guide is currently handled as a working document. New routines are documented before they are fully tested (and occasionally before they are written). The current status of each program is, however, indicated in tables at the beginning of each chapter.

III. Short-Term Goals

We assume that principal input to a VLA post-processing facility is u-v data which have been edited, calibrated, and partially sorted by the Asynchronous Subsystem of the VLA. In order to acquire early experience with such data and in order to assist in the development of the Asynchronous Subsystem and to support users of that system, we feel that our next task is the creation of, at least, a minimal package of programs to deal with Dec-10 output. Thus, during the next 6 months we expect not only to clear up the problems mentioned in the previous section, but also to develop a capability similar to that described above for output from the Asynchronous subsystem.

The requirements of the Asynchronous package are somewhat simpler than those of the Synchronous package. Counterparts of VLATOIMB, VLACLEAN, VLACORR, VLACAL, VLAPCAL, VLAPCORR, and some parts of VLASCRIB will not be needed. Further more, the map programs VMCLEAN, VMINDEX, VMCOPY, VLAPRINT, VLADRAW, and VLAPIX may be used by both packages. Our current personnel (i.e., Eric Greisen and Fred Schwab) should be enough to accomplish these goals in the stated time. A format has been determined and we expect to begin coding the package soon.

IV. Our Perspective on Future Development and Goals

The program packages described in the preceding sections are designed to meet the ongoing requirements of the VLA during its developmental stage. These packages are expected to neither handle very large volumes of data with practical efficiency nor are they powerful or flexible enough to provide substantial improvements over data reduction procedures now in use with older, 2 to 4 element interferometers. The programs which process Synchronous Subsystem output will fall into disuse quite naturally as the Asynchronous Subsystem becomes more highly developed and more reliable. However, programs to handle the output of the Asynchronous Subsystem will need to become more extensive, flexible, and efficient. This latter development is the subject of the present section of this memorandum.

The project outlined below is a limited one which we feel is suited to Phase I in the development of a VLA post-processing capability. Crash projects, involving large numbers of people, or projects which try to solve problems expected to arrive many years in the future, work even less well in software development than in most other areas. The correct approach, we feel, is to develop programs gradually which work on real data to do useful things for real observers. To make the gradual approach acceptable and to allow close interaction between observers and programmers, the project must begin when the number of observers is still quite small. The intent of the project is to develop as much post-processing capability as possible, but not to worry so much whether it can support multiple users and whether it is as efficient as it might be. The proposed system could be viewed as a test system where the emphasis is to give a single user as powerful and versatile a capability as possible. Early work will be limited to continuum reduction, but spectral-line routines might be added when the need arises.

In order to determine what programs are likely to be useful, let us consider the common case of a continuum observer faced with a field containing several sources of interest. The data are first calibrated and edited at the site. Maps, probably at less than maximum resolution, would then be made. At the post-processing facility, the observer begins by inspecting his maps for signs of "bad data" (e.g., ripples or other strange regularities). This inspection may be done most quickly and with substantially better sensitivity using television-type display devices. When questionable areas appear on the maps, the observer would review the calibration data used for the source (and possibly also the u-v plane data taken on the source itself) in order to isolate the bad data. Plots of amplitude and phase versus time for the suspect baselines and possibly gray-scale displays of the u-v plane would be needed. Once any questionable data are located, they might be deleted and new maps made.

The next step could involve fine tuning the standard corrections made to the maps. The observer may wish to examine the effects of correcting for edge attenuation due to the primary beam and to the smoothing and resampling of the FFT. He might decide to correct the maps for the delay beam. Whether or not one makes this correction and exactly how this correction is made depends on the nature of the experiment and the quality of the data. In any case, the astronomer will certainly wish to inspect the result to determine whether he wishes to use the corrected map or a smaller portion of the original map. Single-dish data may now be added to the data to fill in the central portions of the u-v plane. The Westerbork experience indicates that this is a delicate process requiring careful calibration of the single-dish data and close inspection of the results of their inclusion in the maps.

Next the astronomer will need to decide on one or more resolutions and map sizes. While we will probably encourage the use of the smallest FFT which will not degrade the final results, the astronomer may note that a confusing source probably lies just outside his field of view and thus may need a larger transform to isolate it. In order to examine extended features the astronomer will wish to subtract any strong point source from his map (via a CLEAN-like program) and then convolve at least a portion of the resulting map to a lower resolution.

Once the astronomer has finished correcting and massaging his maps, he is ready to derive the useful information from his data and to produce the final displays. In the present multi-source continuum example, he will want to isolate the sources on the map and to estimate their positions, sizes, and flux densities. The first step in this modeling process is to search the map plane for sources and to determine good "initial guesses" for their parameters. For a complex field the astronomer will need to assist the programs, guiding them to the sources and providing them with source-dependent control parameters. When a good set of initial guesses has been determined, a fitting routine, which operates on data in the u-v plane, will be used to refine the parameters. The astronomer will again require various methods to compare the model with the data. Should straightforward model procedures be inappropriate for portions of the map, the astronomer will wish to examine such portions in detail and measure parameters such as the widths in various directions, the peak brightness, etc. For many objects, the next stage is an examination of polarization and spectral index maps. A variety of map presentations are required at this point which include good procedures for blanking regions in which the parameters are unreliable. Finally the astronomer will examine his maps to determine which data are required for documentation and publication of his work and which regions, parameters, and methods to use for published displays of his data. His production of these displays (picture, Calcomp, etc.) end his processing task.

The above operations all involve a detailed interaction between the astronomer and his data involving many correction and parameterization steps followed by examinations of the results. These operations could be accomplished through traditional batch mode processing such as is currently done with the Green Bank interferometer. However, we feel that the long waits required in batch mode for the results of each suboperation lengthen the data reduction process enormously and cause the astronomer to lose his concentration and to take shortcuts in the processing. The latter effects can degrade the final results. The combination of real-time displays with a responsive computer system, similar to that used for single-dish processing at NRAO, would allow the astronomer to process his data more rapidly, to maintain his concentration on those data, to check fully the results of his data manipulation, and to discover, more easily, unexpected problems - or results which may be present in his data. The process of data reduction depends heavily on the nature of the particular observations and on the likes and dislikes of the particular observer. However, experience (e.g., TPOWER/CONDARE) has shown that the presence of a responsive, real-time, data display/reduction system enhances the scientific output and makes efficient use of the computer hardware.

We have alluded in the discussion above to a variety of software routines required for continuum data reduction and analysis. Below we will present a condensed list of the capabilities which should be present, eventually, in the proposed software system. We do not discuss spectral-line problems here because they are somewhat further down the road and because the discussion given in "The VLA Spectral Line System: A Progress Report" (NRAO, January 1976, pages II.B-3 through II.B-15) covers the subject well.

A "wish list" for continuum VLA post processing:

A. Data integrity examination and correction

1. Map routine for large fields with minimal resolution and just adequate accuracy

2. u-v plane display routines including gray-scale (source data) and baseline-time (source or calibrator data)

3. Bad data deletion routine

4. Minor numerical routines for estimation of amplitude and phase variations and noise.

B. Map correction

1. Gridding (e.g., sin x/x)

2. Primary beam

3. Delay beam

4. 3-D smearing

5. Center hole in u-v plane

6. Polarization beam

7. Internal calibration by observed point source response.

C. Map restoration

1. Include data from other telescopes particularly to fill center hole in u-v plane

2. Point source subtraction

3. Clean

4. Maximum entropy on small fields

5. Map combination including fields at different phase centers

D. Convolution of restored maps

1. Gaussian

2. Beam of another radio telescope

3. Extended source enhancement

E. Parameter estimation

1. Position, flux, angular size fitting for simple source components in map plane

2. u-v plane model refining

3. Detailed examination at high resolution of small fields surrounding source components

4. One dimensional fits across extended sources at arbitrary angles

F. Map comparison

1. Polarization

2. Spectral index

3. Rotation measure

G. Special problem software (low priority)

1. Astrometry/proper motion

2. Circular polarization

3. Pulsars

4. Planetary system

H. Display

1. Gray scale (TV and, for hard copy, Dicomed) including labeling, contrast enhancement, pseudo coloring, contour or polarization overlay, et al.

2. Contour and line drawing outputs in both publishable and quick and dirty forms

3. Printer output for processing documentation and detailed results.

Proposed Development System

The capabilities listed above can only be offered with a great deal of work and with a small, but not insignificant, capital expenditure. In the latter area, because of the fundamental nature of map display in the data processing procedure, acquisition of a display system is of high priority.

Experience has shown that unresponsive (i.e., very slow) interactive systems frustrate the user and hence, are ineffective. Thus, when hardware bottlenecks can be identified, corrective hardware (such as an array processor or more core or disk) should be acquired.

We propose to use a combination of the IBM 360/65 and the Modcomp IV both as stand alone computers and as a connected system (at least to the RJE level). Items which involve the Modcomp will normally be made available on a one-user-at-a-time basis, while items involving only the 360 will, of course, be multi-user. The proposed project schedule includes; during the first year, the addition of a graphics capability to the Modcomp (not later than August 1977), with the completion of at least initial versions of roughly half of the above capabilities by mid 1978. In the first year, effort will be concentrated on simply producing the capability. In some cases, its use will be cumbersome because of a lack of a particular piece of hardware or simply because attention was focused on acquiring capabilities rather than making them efficient. In the second year (1978-1979), we expect to refine the algorithms (e.g., to account for 3-D effects and the like), to speed up those that are particularly inefficient, and to offer nearly all of the capabilities listed above. Hardware requirements during the second year will probably include additional core and disk for the Modcomp. During the third and final year of Phase I an array processor would be added to speed up the programs, the algorithms would be extended and refined, and some capabilities peculiar to spectral line processing added. By mid 1980 the post-processing station should be as flexible and as fast as we know how to make it, though it will still have a very limited throughput. An integral part of the entire program will be the close interaction with observers to improve and develop the conceptual basis as well as the tools of map analysis.

The total proposed project cost, to be completed in 1979, is $225K. The immediate step is a microprocessor-based display system ($75K) which should be procurred by June 1977 and installed by August. It is felt that the particular order of items during the following two years should be determined as hardware bottlenecks are found in the use of the system. It is not now clear, for example, whether a little more core or a lot more disk would buy the greater improvement in throughput. The tradeoffs will become clearer when we see which routines are most heavily used and which are the most frustrating.

Future - Phase II

In the 1978-80 time frame NRAO must specify and develop whatever post-processing facility will be used in the 1980's. This system must service the needs of the full, normally operating VLA including both continuum and spectral line observations. Many questions remain unanswered regarding this future facility, not the least of which is the question of where NRAO's basic responsibility ends with regard to the extent of supported data processing. Also unclear is the level of funding NRAO will have at its disposal. These difficult questions are not addressed in this report. It is important to recognize, however, that regardless of what level of post-processing the NRAO decides to support, the two or three year program outlined here can be viewed as a necessary prerequisite. The kind of facility in need will not evolve by simply signing a contract with IBM or the like. Rather, at least several man years of difficult work is first required. It is that endeavor which is proposed here.