VLB ARRAY MEMO No. 371

National Radio Astronomy Observatory
Edgemont Road, Charlottesville, VA 22901
(804)296-0211, FTS=938-1271

18 July 1984

To: VLBA Post Processing Working Group

From: Bill Cotton

Subj: Thoughts on VLBA-VLA Editing and Calibration and AIPS.

1.0 INTRODUCTION

This memo contains my current collected thoughts about implementing a set of synthesis telescope data editing and calibration routines into the AIPS data reduction package. I will be maintaining a wish list for calibration and editing and invite all comments and suggestions.

2.0 TO AIPS OR NOT TO AIPS

The first point to be considered is the question of whether AIPS is the appropriate context for extensive editing and calibration development. The traditional use of AIPS has been for processing data which is relatively well edited and calibrated so that the tools available for these purposes are rather crude. A more fundamental problem is that the manner in which data are stored and manipulated is not well suited to editing and calibrating large amounts of data on many sources. Currently, calibrated data are kept as one source/frequency band per catalogued file, and operations on data generally result in a new copy of the data. This is not well suited to editing and calibration in which many, relatively simple operations are done on the data.

There are several reasons in favor of implementing editing and calibration in AIPS. The first is that it is much simpler to make the necessary modifications to AIPS than the develop an entirely new system from scratch. A second, much more fundamental, reason is that the eventual users of this system will not wish, and should not be required to learn yet another system.

A further reason is that this calibration software would greatly strengthen AIPS as a general purpose package. Given these considerations, I believe that implementing calibration and editing in AIPS is appropriate.

3.0 DATA STORAGE.

3.1 Long Verses Short Term

There are actually two types of data storage to be considered, archival or long term storage on tape and catalogued disk data files. (Individual tasks may have their own internal format for scratch files but that is a rather minor issue.) Since we have decided to adopt FITS as our standard tape format we need to take care not to violate the international standards for FITS. The internal format in AIPS is not, in principle, subject to external standards but since it is patterened after the FITS format the questions of external and internal data storage are coupled.

There are several aspects of the current DEC10 approach to storing data which are quite attractive for purposes of editing and calibration. These are: keeping all sources in an observing session in a single file, and maintaining the data in a raw form with an associated gain/calibration table. This latter aspect of the VLA DEC10 package allows the user to start over fairly easily when he decides that he has the data calibration hopelessly fouled up (frequent with novice users) yet doesn't fill up the disk as the current AIPS approach is prone to do. I should note that the raw data with gain table approach after calibration is completed becomes very expensive, both in CPU time and programming effort; the difficulties with the VLA pipeline illustrate this point.

3.2 Optical Disks

The optical disk technology appears to be at the point that, within a few years or less, optical disks will be in common use for archival storage and will likely be used for both the VLA and VLBA. This archival storage could be either FITS or AIPS internal format. FITS is preferable as an archive format since a given FITS file should be readable indefinitely but the AIPS (or any other program package) internal format is subject to occasional revision. For this reason to archival storage should probably be FITS.

If the optical disk storage becomes sufficiently cheap then it may be desirable to keep a temporary version of the data in AIPS internal format on an optical disk. Since these are primarily write-once devices, the data file and calibration files would have to be on separate disks. Any flagging by the user would also need to be kept in a magnetic disk file rather than with the data. When optical disks become available, AIPS should be modified to allow separate files of a catalogue entry to be split between different storage devices.

3.3 Multi Source Files.

As mentioned above, the current AIPS (and FITS) files contain data for a single source and observing band. This is relatively convenient for edited and calibrated data which are then processed for each source/band separately but is not convenient for editing and calibration, during which data for many sources may be processed together. Current VLA observing programs can contain as many as several hundred separate sources and the VLBA may produce data on similar numbers of sources. Under these conditions individual source files are extremely cumbersome.

Although current AIPS catalogued files and FITS tape files contain data on only one source, generalization to multi-source files is mostly a philosophical exercise. Both FITS and AIPS formats allow for a relatively arbitrary number of "random parameters" associated with each data record and both allow for a wide variety of associated tables. Multisource files could be implemented by adding a source label random parameter to each data record and a table with the source dependent information. Separate observing bands (e.g. 6 and 2 cm) are normally processed independently so that there is no compelling reason to have multi band files.

3.4 Gain Tables.

The current practice in AIPS is to keep fully calibrated data and, when the data is modified, to create a new data file. This works fairly well when the operations on a data set are few and fairly major but can be very expensive in disk space during editing and calibration when the files can be very large and the operations fairly numerous. For this reason gain tables may be necessary for the editing and calibration phases (although I am opposed to their use during later processing).

AIPS currently has a gain table which is produced by ASCAL and its clones. This is a gain table which has been applied to the data and can be applied to another data set. In principle, data sets could have similar gain tables which have not been applied. This would allow maintaining files with the raw data but with gain tables which could be modified at will. This approach requires data display routines which can apply the gain tables but these need to be written anyway.

The principal difficulty with having gain tables associated with data is that it is difficult to sort the data. We may be able to live with data in only time-baseline order for the calibration and editing phase. Since gain tables are relatively short this problem may not be especially serious. If data are stored on optical disks they cannot be sorted without making a new copy of the data set.

3.5 Software Necessary

Since the approach outlined above introduces new types of files into AIPS some amount of new utility software is needed. AIPS currently has two types of catalogued data files, maps and uv data. Multi source files with gain tables could either be uv type files or a new type. I don't see strong arguments either way but it is probably cleaner to invent a new type. In either case, it will be necessary to have a utility task to take a multi-source, raw data file and to apply the calibration and to split the data into separate source data files so that existing data processing software can be used.

4.0 EDITING

4.1 Data Display

The principal problem of editing data is the proper display of the data to allow the user to see problems easily. The traditional VLB approach of plotting the data on each baseline and allowing the user to flag data one record at a time is too cumbersome for large data sets (but some users may insist on it).

One approach, which Jim Torson worked on at the VLA and we have discussed in Charlottesville, is to convert the data into an image which can be displayed on the television. The data amplitudes and/or phases could be gridded with, say, one axis being time and the other being baseline. An interactive flagging program could then allow the user to specify data to be flagged by specifying a region of the image. This scheme probably only works well if a single source is given in each display.

4.2 Reversibility

In general, anything a user can do to his data he should be able to undo. This is especially true of editing. Currently in AIPS there are two levels of flagging (although we don't take full advantage of this). Each visibility value has an associated weight, and non positive weights are taken to indicate flagged data. In a sense this gives two types of flags: reversible (negative weight) and irreversible (zero weight). Data flagged on the basis of monitor data or by the correlator could be given zero weights and data flagged by the user could simply have the sign of the weight made negative. This scheme may be sufficient but more thought should be given to the subject.

4.3 Software

There are a large number of potentially necessary and/or useful display and editing routines; a number of the more obvious of these are listed below:

1. Gridding - This routine would take uv data and create an image which could be displayed on the television. The axes of the image could be relatively arbitrary but some possibilities are time-baseline and u-v.

2. Interactive Flagger - This routine could take the result from the gridding routine described above and allow the user to interactively flag data.

3. Traditional Baseline Display/Flagger - This routine could display two dimensional plots on a graphics device of amplitude, phase, delay etc. as a function of time and allow the user to interactively flag data.

4. Printer Data Display - This routine is like the VLA DEC10 LISTER program which would allow the user to print data in a large variety of different forms.

5. General Flagging - This routine would be an enhanced version of the AIPS routine UVFLG similar to the DEC10 FLAGER. This routine would allow the user to specify the data to be flagged or unflagged.

5.0 CALIBRATION

In the following discussion I will be addressing primarily the problems of calibrating continuum data as I am not personally familiar with the detailed problems encountered in calibrating spectral line data. I will also not address polarization calibration as a polarization calibration task is being developed in AIPS by Larry Molnar at SAO.

The basic elements necessary for total intensity, continuum calibration currently exist in AIPS with the tasks ASCAL and ASCOR which will determine antenna gains from data and a source model (ASCAL) and apply a gain table to a second data file (ASCOR). The existing software probably needs work especially in the area of multi frequency data sets which are becoming common for the VLA with the A, B, C, and D IFs and will probably the standard mode for VLBA data.

Polarization calibration software is currently being developed by Larry Molnar at SAO. This AIPS software should be usable for both VLA and VLBA data.

5.1 Geometric Observables

The VLBA processing software will have to maintain accountability for the geometric model which has been applied to the data. This requirement will have an at least indirect effect on the calibration software. Our current thinking on this subject is that each data record will have random parameters giving the total model group and phase delay and their time derivatives. Since the calibration routines will effectively change the model values of these, corresponding changes need to be made to the appropriate random parameters.

5.2 Atmospheric Corrections

Accurate weather/atmospheric corrections will need to be applied to the data. Many of these will be done in the fringe processor for the VLBA and some of this may be done on the VLA. The ability to correct or improve the atmospheric model is fairly important. These models may involve monitor data such as water vapor radiometer data, Faraday rotation monitor data, ionosonde data, weather bureau data, temperature data, humidity data etc.

5.3 Multifrequency Data

There are two separate problems with multi frequency data sets. The first is when there are several receivers involved such as the VLA using all four IFs. In this case, a separate solution must be found for each IF.

The second case is when multiple frequency channels are derived from the same IFs as for the VLA in the spectral line mode or the VLBA for all data. In this case a single solution is needed for each IF but, for the VLBA, the group delay may be sufficiently poorly determined that there is a phase slope across the frequency domain. Removing this frequency phase slope is commonly called fringe fitting and could be incorporated into the standard calibration routines. Fringe fitting is currently done separately in the routine VBFIT and its clones.

5.4 Calibrator Source Models

The current calibration practice with the VLA is to assume that the calibration sources are point sources at least for some range of baseline lengths. This is not completely satisfactory for the VLA since the two primary flux density calibrators, 3C 48 and 3C 286, are both seriously resolved in some configurations at some frequencies. For the VLBA there are probably no sources which can be used as point calibrators.

Thus, it appears necessary to use more elaborate source models for the calibrators. These models could be either provided by the observatory or derived directly from the users data by the usual self calibration techniques.

The net effect of this is that it will probably be necessary to work on each calibrator source independently and then merge the gain tables from different calibrators.

5.5 Software

Most of the necessary software changes are enhancements of current capabilities. In addition, the current format of the gain table is not sufficiently general and probably needs to be redesigned. Another problem is the location of the gain table; If there is a single data file with all sources the gain table can be associated with this catalogued file. Several areas requiring work are listed below:

1. Multifrequency ASCAL - This enhanced version of ASCAL should be able to handle at least 4 independent IF and should be able to average a number of frequency channels derived from the same IF. It may be necessary or desirable to incorporate fringe fitting capabilities into this routine.

2. Polarization Calibration - This is being worked on by Larry Molnar at SAO.

3. Gain Table Manipulation - There needs to be a routine or routines to manipulate gain tables. Some of these functions are: concatenating gain tables, editing gain tables, displaying gain tables, and smoothing gain tables.

6.0 GEOMETRIC CALIBRATION

By geometric calibration I mean the correction of the data for the effects of the geometric configuration of the array. In the past these effects have been mostly of interest to astrometrists and geodesists but are likely to become important to astronomers also. The list of these effects seems infinite but a few which we will most likely have to face include: polar motion, fluctuations in the rate of UT1, solid earth tides and ocean loading. In addition, we will probably/certainly need the capability to correct the correlator models as better source and antenna positions become available. Several general categories of software yet to be written:

1. Detailed Modeling - The models for polar motion, UT1 etc. etc, and etc. need to be applied to correct the phase of the data. Many of the models will be determined externally but some, such as UT1 corrections, may be determined from VLBA data.

2. Model Correction - Corrections or improvements to the previous geometric model should be possible.