VLB ARRAY MEMO No. 400

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

17 October 84

To: VLBA
From: W. D. Cotton
Subject: Data Processing Meeting 16 October 1984

Present: Benson, Burns, Cornwell, Cotton, Fickling, Greisen, L. Molnar, Romney, Schwab, R. Simon, Walker, Wells

C. Walker complained about the amount of premeeting material being sent by mail. In the future this material will be put in a file in CVAX::UMA3:[VLBA.DATAPROC] with the name of the file being the day of the meeting (e.g. 16OCT84.TXT). E. Greisen requested that an abstract of the material be distributed before the meeting.

R. Simon reported that the new deliverly date for the memo on the requirments for geometric work with the VLBA which K. Johnston promised to write is 1 December.

The focus of the meeting was a memo by B. Cotton outlining the techniques for handling uncalibrated data in AIPS which have been developing over the last few months. An updated version of this memo is appended to the end of this document.

There was a general agreement that flagging needs to be possible on time scales shorter than the interval between gain table entries and for various reasons the flags associated with the raw data should not be changed. Since a complete table of flagging information with the time resolution would require much disk space and be mostly empty some scheme of keeping a list of user flags was favored. The details of this implementation were not completely worked out.

In the discussion of flagging, J. Romney expressed the view that flagging at both the prior (on line) level and the user level should be multilevel. The current use of the sign of the weight in the allows two levels of flagging, reversible (negative weight) and irreversible (zero weight). No one suggested that this would be inadequate.

T. Cornwell suggested using RMS instead of the current weights and thought that it is possible to get this information for the VLA. This approach appeared desirable but the details of the implementation were not settled. The principle problem is that in the current use of weights, an increasing value increases the effect of the datum; whereas, if the RMS were used instead, the effect of a datum should be inversely proportional to the RMS.

There was a discussion of the definition of scan to be used to create the index file. It was generally thought that the current meaning of scan was inadequate and, in the case of long sequences of data on the same source, the tabulation in the index file should be more frequent than every source change. The maximum length of a "scan" ,for purposes of the index file, should be either scaled to the amount of data or user controlled.

C. Walker volunteered himself and J. Benson to determine how to handle the geometric observables in the data base.

The current plans for the software interface to the database is to have standard access routines with a standard set of selection criteria. Most of the discussion in this area concerned the data selection adverbs for the user interface. L. Molnar and T. Cornwell supported the use of source qualifier as a standard selection criterion.

There was considerable discussion about selecting multiple sets of timeranges, frequency ranges etc. There was some sentiment for making the corresponding adverbs arrays to contain such information. C. Walker suggested that complex data selection could be done by developing a file of selection criteria. This could be either an explicit selection file or a temporary set of flags to deselect unwanted data.

There was an extended discussion about the nomenclature and specification of multiple frequency bands. (This is the problem of the A and B and C and D ifs on the VLA and the data derived from different video converters on the VLBA) It was generally considered that this problem was sufficiently widespread that the concept should be introduced into the user interface. For current purposes we will define an "if" to consist of a set of uniformly spaced frequency channels (possibly a single channel) and all polarizations for those frequency channels. This may not be a satisfactory terminology since it differs from the current usage at the VLA. In any case, it was agreed that the user should be able to select data by "if".

There was a general consensus that the data selection AIPS adverbs should be able to specify "all except..." wherever practical.

T. Cornwell suggested that the selection criteria should be able to specify data by baseline including the form "antenna i with all". He also suggested that the time range be specified by giving civil day notation instead of a day number; there was considerable dissent on this issue and little agreement as to the proper method to specify day.

APPENDIX

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

To: VLBA
From: W. D. Cotton
Subject: Raw UV Data Files in AIPS.

1.0 INTRODUCTION

The purpose of this memo is to outline a suggested scheme for dealing with raw (uncalibrated and unedited) uv data in AIPS at the system level. That is, to describe the information to be kept and the software interfaces to the data. There will be no attempt to specify the specific calibration/editing functions. The scheme proposed is very similar to that currently used on the VLA DEC-10.

The basic proposed scheme is to keep the data in the raw form in multi-source files in time order. This data, including prior flagging information, would never be modified but a gain table would contain the information necessary to flag and calibrate the raw data. The gain table will contain information which varies on fairly short time scales (seconds to minutes). Calibration information which is much more slowly varying, such as bandpass calibration, will be kept in separate files.

Flagging information will be kept in a file which contains a list of flagging criteria. There may be several sets of these files to be used in selecting data.

In order to perform an editing or calibration function, a task will call a subroutine which will select data by given criteria and optionally apply calibration and/or user or prior flagging information. Calibration and/or editing information is then sent to an interface routine(s) which will update the gain table and other tables as appropriate. Thus access to the data set will be thru a single set of interface routines

For many purposes, a subset of the data may be selected and converted to a more traditional AIPS uv or image file for extended processing. An example of this is when a calibrator source needs to have several iterations of self-calibration in order to derive a proper model of the source. In this case, the calibration information derived in this process would be used to update the gain table.

The following sections will outline the suggested data base structure and the software interface to the data base. Many of these items will be defined in more detail in the appropriate specification

files.

2.0 UV DATA BASE FOR RAW DATA.

The form in which data will be kept is raw data with associated on-line or other prior flagging information with all sources at a given frequency band in a given experiment being kept in a single, time-ordered file. Associated with this file are files which contain information necessary to reject bad data and to calibrate the data. All modifications are made to the calibration (gain) and editing (flag) tables.

2.1 Multi Source Files.

Each visibility record will contain a random parameter which is its source number. This number will indicate the entry number in an associated table extension file which gives the name, position and other information about the source.

2.2 Frequency Channels And Ifs.

The data base will contain separate axes for frequency channels and ifs. For the present purpose, an if is defined to consist of one or more equally spaced frequency channels and all polarizations observed in those frequency channels. Examples of this are the A and B receivers of the VLA the any of the 16 video converters of the VLBA. Ifs may be irregularly spaced. This definition allows for several irregularly spaced sets of regular frequency channels.

2.3 Time Order

The records in the file will be in time order. The time tags of the records will be nondecreasing irregardless of the source. This precludes sorting the file but allows indexing the file.

2.4 Flagging

There will be two kinds of flags associated with the data. The first kind, referred to hereafter as prior flagging, is flagging information applied to the data before it is initially loaded into AIPS. This prior flagging will usually be online flagging produced either by the antenna monitoring system or at processing time for VLBI data and is kept in the visibility records. Prior flagging is never changed but may be entirely ignored if so specified by the user.

The second kind of flagging, referred to as user flagging, is flagging information in a flag table(s) which results from user action. These flags may be cleared or modified by the user. The flag table will consist of a list of specifications of data to be rejected (i.e. timerange, antennas or baselines, polarizations, ifs, or frequency channels).

2.5 Gain And Other Calibration Tables

The raw uv data file will have an associated file or files which contain calibration and editing information. The average time spacing of the entries in this table will be specified by the user but will be adjusted such that each entry in a given scan covers the same or nearly the same time range. Calibration information which may vary on time scales of seconds to minutes are contained in the gain table. Information which is more slowly varying such as bandpass calibrations will be kept in separate tables.

The gain table will be created and initialized at the time the data file is filled and will at all times contain the current calibration information.

2.6 Index File

Each raw uv data file will have a file which is created and filled at the time that the main data file is filled which contains an index of the data in the file. There will be an entry for each scan which gives the source, timerange, and other necessary information as well as the range of visibility numbers corresponding to that scan. A scan is defined here to consist of a sequence of data which terminate when the source or observing band changes or there is a time gap in the data which is much longer than the integration period. When this definition allows long sequences of data to be described by a single index entry a maximum length may be imposed on a "scan" and a given sequence of data will be broken into several index entries.

2.7 Geometric Observables

The geometric observables (total model delay, rate etc.) will be kept in the raw data as random parameters. Derived values (from fringe fitting etc.) will be kept in the gain table. Thus, when processing is finished the gain table should contain the total model values of the geometric observables which will then be the observed values.

3.0 ACCESS TO THE DATA BASE

Access to the data base will be thru a set of interface routines. This will allow for future modifications to the structure of the uv data base which may be transparent to the applications software. In addition, there will be a standard set of data selection criteria which can be set by the user to specify the data to be selected and the calibration to be done.

3.1 Data Selection Adverbs

The standard set of selection criteria will be defined by a list of AIPS adverbs. Other selection criteria which are peculiar to a given application must be applied by the appropriate application software. A list of these adverbs and a short explanation follows:

- SOURCES: an array of source names to be selected or deselected.
- QUAL: source qualifier to be selected.
- CALSOUR: an array of calibrator source names to be selected or deselected.
- TIMERANGE: a time range allowed for data.
- UVRANGE: the range of baseline lengths to be selected or deselected.
- STOKES: the stokes parameters of the data selected or deselected
- IFS: the if to select or deselect.
- BCHAN: first channel number in the specified if to be selected or deselected.
- ECHAN: last channel number in the specified if to be selected or deselected.
- ANT1: an array specifying the antennas to be selected or deselected.
- ANT2: an array of antennas to be used with the antennas specified by ANT1; baselines involving antennas explicitly or implicitly specified in ANT1 with antennas explicitly or implicitly specified in ANT2 are selected.
- SUBARRAY: the desired subarray.
- DOCALIB: If true, calibrate the data.

- INTERPOL: specifies the interpolation method to be used to obtain calibration from the gain table.
- INTTIME: the time scale of the gain smoothing.
- PASSFLAG: specifies whether to use the prior flagging, the user specified flagging or both.

3.2 UVGET

UVGET is the routine which will read the raw uv data file apply specified calibration and selection criteria and return the visibility data to the calling applications task. This routine will make use of the index file to speedup reading small amounts of the data base. This routine should be the only read access to the raw uv data.

3.3 CALPUT

CALPUT will be the interface to the gain table. Applications tasks wishing to modify the information in the gain table will call this routine. Calling routines will pass a derived gain table to be used to modify the raw uv file gain table. Other routines may be necessary to modify the more slowly varying information in the other calibration files.