Very Long Baseline Interferometry and the VLBA
ASP Conference Series, Vol. 82, 1995
J. A. Zensus, P. J. Diamond, and P. J. Napier (eds.)

Printed June 13, 1995

CHAPTER 12

VLBI Data Reduction in Practice

P. J. DIAMOND

NATIONAL RADIO ASTRONOMY OBSERVATORY

Abstract Practical guidelines to data processing in AIPS are given.

Contents

12.1 Introduction

Other chapters in this book have dealt with the theory of Very Long Baseline Interferometry or have touched upon the data processing in a relatively abstract fashion. The purpose of this chapter is to describe the steps that a typical user of the VLBA might follow when reducing a dataset. The aim is to provide a description of the basic steps that a user should follow in order to produce VLBI data that is ready for self-calibration and imaging. The examples shown are those available in the 15JAN95 version of the NRAO AIPS package. This chapter is complementary to the VLBI chapter of the *AIPS Cookbook.*

12.2 Calibration Philosophy

The philosophy adopted by the AIPS system with regard to calibration is summarized in one phrase: *data are sacrosanct.* By that we mean that during the calibration process the raw data are never touched, all calibration information is stored in tables attached to the data files. These tables can be manipulated in many different ways, they can be merged, smoothed, interpolated, extrapolated, even deleted. The AIPS software can apply existing calibration tables on-the-fly so that the user can see the effects of a particular portion of the calibration upon his or her data. After a user is happy with the calibration the appropriate tables are then applied to the data producing an output file of calibrated data that is then ready for imaging and/or further self-calibration.

The calibration procedure is almost exclusively antenna-based. There are, of course, some effects that need to be dealt with in a baseline based manner and such calibration can be accommodated within the AIPS system.

All (u, v) data within AIPS are stored in the frequency domain and all calibration operations are performed in that domain. If the data are required in the delay (or lag) domain (as is often the case for particular types of operation) then the Fourier transform is performed in memory and the raw data are never modified.

12.3 AIPS Calibration Tables and Data Structure

Before describing the steps a user might follow it is useful to take a closer look at the AIPS table types and their contents and also the structure of a typical dataset within AIPS.

An AIPS (u, v) data file is described by a header such as:

```
Image=MULTI     (UV)         Filename=CYGA        .RAW   .   1
Telescope=VLBI               Receiver=MKIII
Observer=                    User #= 1046
Observ. date=09-JUN-1991     Map date=17-JUL-1994
# visibilities    242180     Sort order  TB
Rand axes: UU-L-SIN  VV-L-SIN  WW-L-SIN  TIME1  BASELINE
           SOURCE  FREQSEL  INTTIM  WEIGHT  SCALE
----------------------------------------------------------------
Type    Pixels   Coord value  at Pixel    Coord incr   Rotat
COMPLEX      1    1.0000000E+00    1.00 1.0000000E+00    0.00
FREQ         8    4.9732400E+09    1.00 5.0000000E+05    0.00
STOKES       1   -2.0000000E+00    1.00-1.0000000E+00    0.00
IF           7    1.0000000E+00    1.00 1.0000000E+00    0.00
RA           1     00 00 00.000    1.00       3600.000    0.00
DEC          1     00 00 00.000    1.00       3600.000    0.00
```

```
--------------------------------------------------------------
Coordinate equinox 2000.00
Maximum version number of extension files of type FQ is   1
Maximum version number of extension files of type HI is   1
Maximum version number of extension files of type AN is   1
Maximum version number of extension files of type NX is   1
Maximum version number of extension files of type CL is   1
Maximum version number of extension files of type SU is   1
Maximum version number of extension files of type FG is   1
```

A glance at the header tells the user that the observations were taken on 1991 June 9. There are 242810 visibility spectra in the (u, v) file. Looking at the FREQ row the user can see that there are eight spectral channels within each IF *(Intermediate Frequency Channel),* each of width 0.5 MHz. The file has seven IFs and only one Stokes parameter (i.e., polarization). Below the main header is a list of the tables attached to the file.

There are a few fundamental tables that the calibration system requires to be attached to each data file. These are:

- **AN:** antenna table, contains information on the geometry of the array being used, the antenna names, their numbers, types (e.g., alt-azimuth or equatorial mounts), coordinates and the time system of the observation (UTC or IAT).
- **FQ:** frequency table, contains the frequency offsets from the header reference frequency for each IF, the bandwidths of each IF etc.
- **HI:** history table, lists what you did to your data.
- **SU:** source table, contains descriptive information about each source, its name, position, velocity, flux density etc.
- **NX:** index table, contains various pieces of information about each scan (or user-defined time interval). This table is useful in providing a data summary and is used extensively by the AIPS IO system.
- **CL:** calibration table, contains all the current calibration information that the user wishes to apply to the data, e.g., amplitude, phase, delay and rate corrections.

Other tables that might be produced by various tasks during the calibration process are listed below:

- **SN:** solution table, contains solutions from an AIPS task—e.g., fringe-fit, self-cal or autocorrelation fitting solutions. Can be applied to the **CL** table.

- **BS:** baseline-solution table, contains baseline-based fringe-fit solutions from BLING. Can be applied to the **CL** table.
- **BP:** bandpass table, contains bandpass correction functions to be applied to the data.
- **FG:** flag table, contains data flagging information.
- **BL:** baseline table, contains additive and multiplicative correction factors that can be applied to the data.
- **TY:** Tsys table, contains Tsys and Tant information for each antenna.
- **HF:** Haystack FRNGE table, contains results from the Mark III Haystack FRNGE program, these are passed by MK3IN. Also can be created by CL2HF within AIPS.
- **IM:** Interferometer Model table, contains the polynomials used in the VLBA correlator to generate the geometric model.

Other tables may also be attached to the data file. A full list of possible tables is given in the AIPS VLBI Cookbook chapter.

12.4 The Data Processing Path

The aim of this chapter is to lead the user through the necessarily complex calibration process necessary to generate VLBI data that is ready for imaging. Figure 12.1 shows an overview of the whole process in block diagram form. As can be seen there are not a large number of basic steps to be performed but, due to very nature of VLBI, the user has to be aware of a lot of the details of the procedure. Hopefully as the VLBA matures a lot of these details can be hidden and VLBI can be made more accessible to the non black-belt VLBI'er. In what follows I shall assume that the data were correlated on either the VLBA correlator or a Mark III correlator. Mark II VLBI is now obselete, may it rest in peace.

12.4.1 Reading the Data

Figure 12.2 shows the steps that must be followed when reading data from either the VLBA or a Mark III correlator. Since the VLBA correlator can process up to 20 stations at a time (see chapter 7) it can write all baselines simultaneously, in the correct time order, to an output FITS binary file. The VLBA tape distribution system then appends a series of auxiliary tables to the data and writes a FITS format tape. The AIPS task FITLD can then read many of these tapes sequentially extracting the data in the correct form for further AIPS processing. Unfortunately, Mark III correlators can only rarely correlate a whole observation in one pass and even then the data format written to tape is a baseline-based

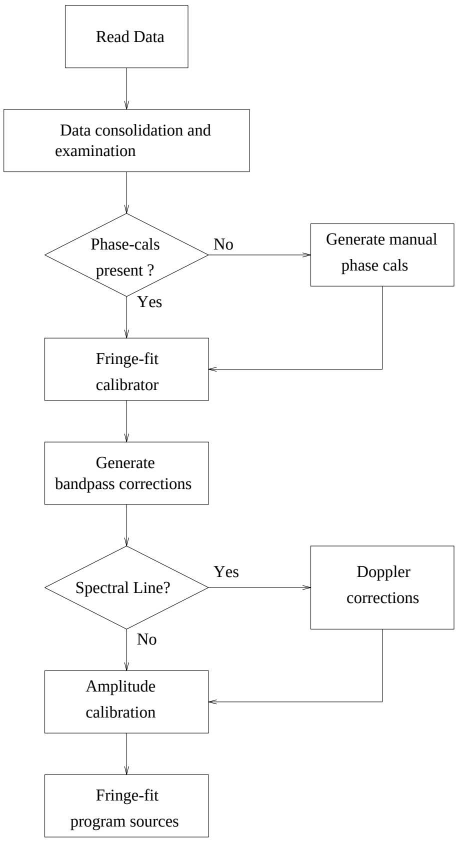

Figure 12.1: Block diagram of the basic data processing steps.

format. Although this format is dealt with in an efficient manner by the software written by the Haystack group, some work is needed to concatenate, sort and merge the data into a form in AIPS that is then identical to that produced by FITLD.

12.4.2 Data Examination

The data used as the principal example in this chapter consist of a 12 station Mark III mode B observation of the core of Cygnus A. The data were taken on June 9, 1991 at a wavelength of 6cm and correlated at Haystack. The observation ran from 18:00 on June 9 to 16:00 on June 10. 12 antennas participated although for technical reasons two antennas showed no fringes and were flagged in the initial stages of data processing.

One of the first things that a user should do is to examine the data file as a whole. AIPS provides several tools for data examination of various kinds. Users

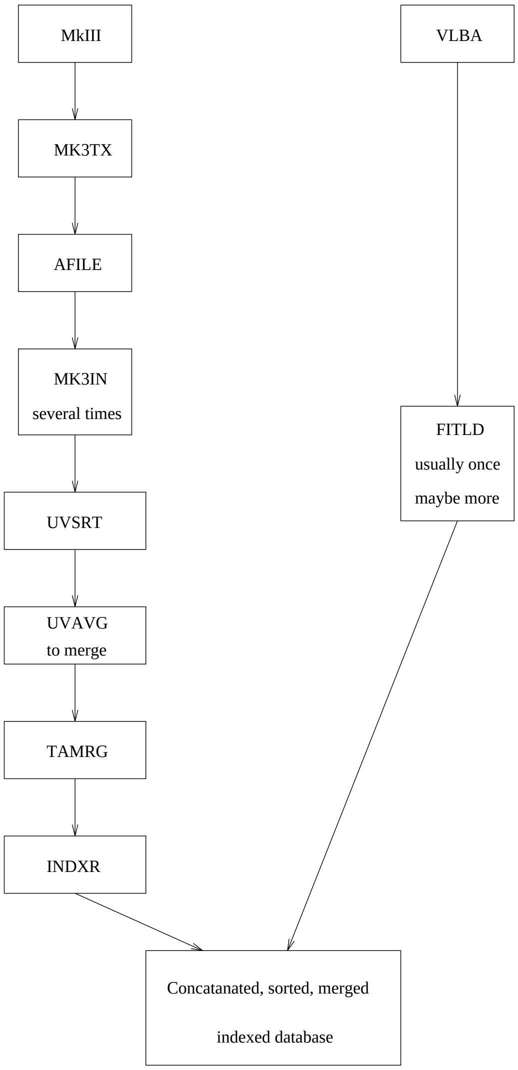

Figure 12.2: A block diagram of the procedure required to read data into AIPS from a Mark III correlator (left-hand side) and the VLBA correlator (right-hand side). As can be seen, for Mark III data a large number of individual AIPS tasks must be run in order to produce an AIPS (u, v) data file ready for further processing.

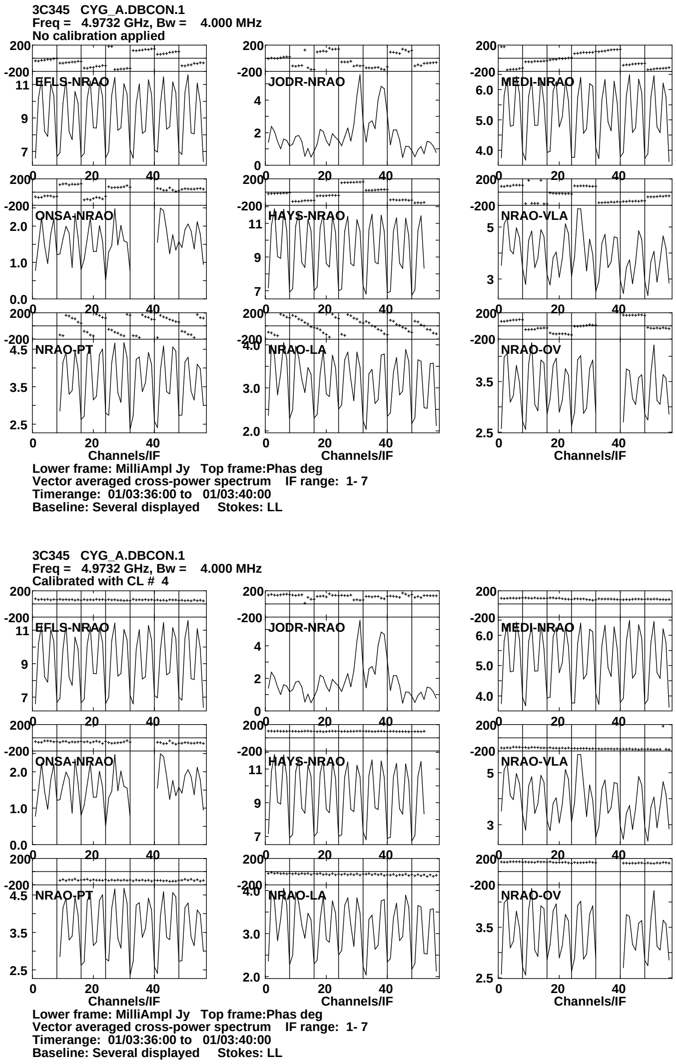

Figure 12.3: The upper plot shows the spectra averaged over a time span of 4 minutes on all baselines to the NRAO 140-foot for the calibrator source 3C345. No calibration has been applied to this data, these are the raw spectra produced by the correlator. The lower plot shows the same data after determining and applying the phase-cals, see section 12.4.3. The phase-cals were determined on a different scan than the one shown so the small residual phase-slopes with frequency are a result of the variation of the multi-band delay with time. Multi-band fringe-fitting (section 12.4.6) will remove such slopes.

can obtain printed listings by using any of the following tasks:

- DTSUM: provides a global summary in matrix form. The matrix lists the number of visibilities per baseline for either the whole observation or on a scan by scan basis. DTSUM also provides a list of antenna numbers and names and the integration times of the data.
- LISTR: provides a listing of the index table, this is convenient for determining the time ranges over which individual sources were observed, the number of visibilities in a given scan etc.
- PRTAN: provides a formatted listing of the antenna table.
- PRTUV: allows the user to print the raw visibilities in a simple tabular form.
- PRTAB: provides a listing of the raw values of any table associated with a (u, v) data file.

Complementary to these tasks are the AIPS tools that allow users to display their data in a graphical format. AIPS offers several different methods of displaying data, some of the more widely used tasks are:

- POSSM: provides plots of spectra in a variety of ways.
- VPLOT: provides visibility plots against time. VPLOT can plot amplitudes, phases, weights, source numbers, integration times etc.
- CLPLT: provides plots of closure phase against time
- UVPLT: similar to VPLOT but does not easily separate baselines into separate plot frames.
- SNPLT: provides plots of various values from SN, CL or TY tables. SNPLT can plot the selected data against time, antenna elevation or hour angle.

Figure 12.3 shows a plot of the raw, uncalibrated spectra for all baselines to the NRAO 140-foot. The vertical lines within each plot delineate the IF boundaries. As can be clearly seen there are many problems with the data at this stage, the phases are not flat across the band in general, they do not line up from IF to IF and the amplitudes are not similar across the band. This is typical of totally raw VLBI data. At this stage any gross editing required should be done.

12.4.3 Lining up the Phases

The first data processing task that users normally tackle is that of lining up the phases between IFs. If the observation was performed with the phase-cal system on at each antenna, and the data being processed was from a Mark III

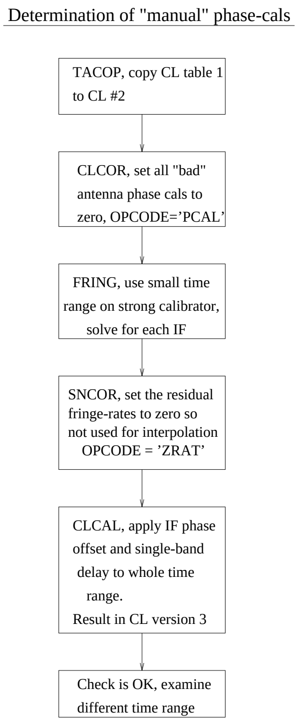

Figure 12.4: A block diagram of the procedure normally used in AIPS to generate manual phase-cal information.

correlator, then this is simply a matter of applying the calibration (CL) table written by the AIPS task MK3IN. Often however some antennas do not have correctly functioning phase-cals for one reason or another and so users may have to generate "manual phase-cals" to line up the data from IF to IF. Figure 12.4 shows the procedure used to generate the phase-cal values and store them in the CL table. The lower frame of figure 12.3 shows the result of applying these phase-cals to the same data as shown in the upper frame. Note that the phase-cals were actually determined on a different scan on the calibrator source 3C345. It is useful to examine the effects of applying the phase-cals to a scan other than the one for which they were determined in order to verify that the correct values are being used. In the near future the phase-cals generated by the VLBA pulse-cal system will be available within AIPS. The use of these data will simplify the process of IF phase alignment.

12.4.4 Bandpass Calibration

Continuum users have not traditionally performed bandpass calibration. This is partially historical in that older fringe-fitting software would determine the

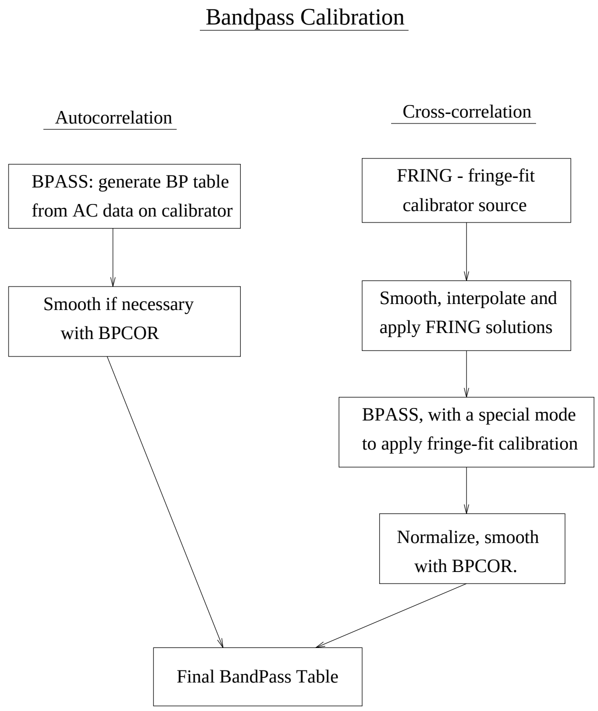

Figure 12.5: A block diagram of the procedure normally used in AIPS to generate a bandpass calibration table.

amplitude and phase in the lag domain and write out these numbers for further processing. Even in software packages that kept data in the frequency domain continuum users would often ignore the capabilities available to line users with regard to bandpass calibration. This has, at some level, limited the dynamic range attainable since a straight average across the band will, in general, cause the amplitudes to be biased lower than their true value. The procedure that users should follow is outlined in figure 12.5.

The plots shown in figure 12.6 are all autocorrelation spectra. Such data can be used to correct cross-power data for the effects of the bandpass filters (see chapter 11), however with this method no correction can be applied for any phase effects. The cross-power data on a strong, well-detected continuum source can be used to correct for any phase ripples in the bandpass. The right hand side of figure 12.5 describes the methodology one would use.

12.4.5 Amplitude Calibration

The theory of amplitude calibration is described in some detail in chapters 9 and 11. The practical aspects are fairly simple but, due to the nature of VLBI,

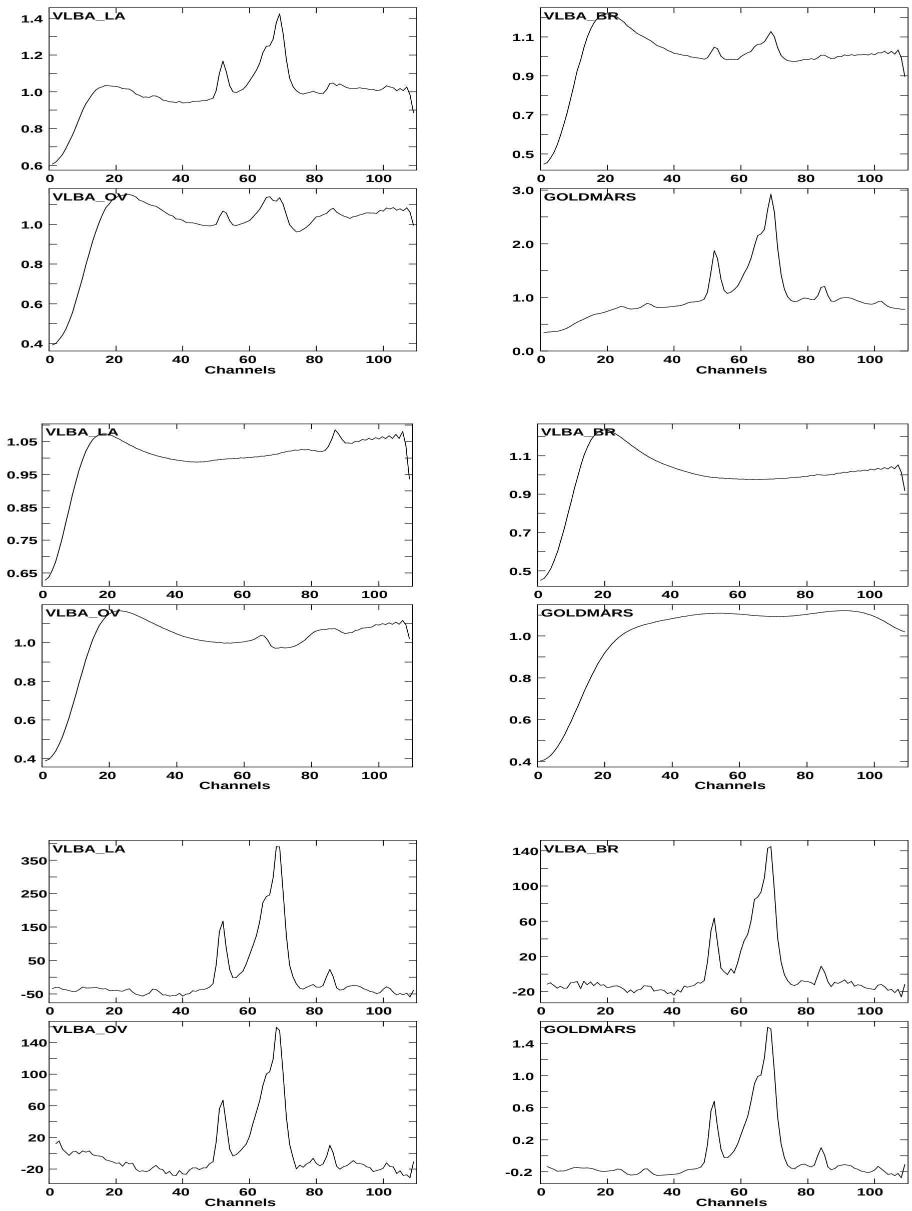

Figure 12.6: The upper four plot frames show the raw total-power spectra of the 22 GHz H_2O masers in Sgr B2N. Note the effects of the bandpass filters. The middle four plots show the spectra on the continuum calibrator source NRAO 530. These autocorrelation spectra were used to correct the maser spectra for the effects of the bandpass filters. The lower four plots show the corrected spectra.

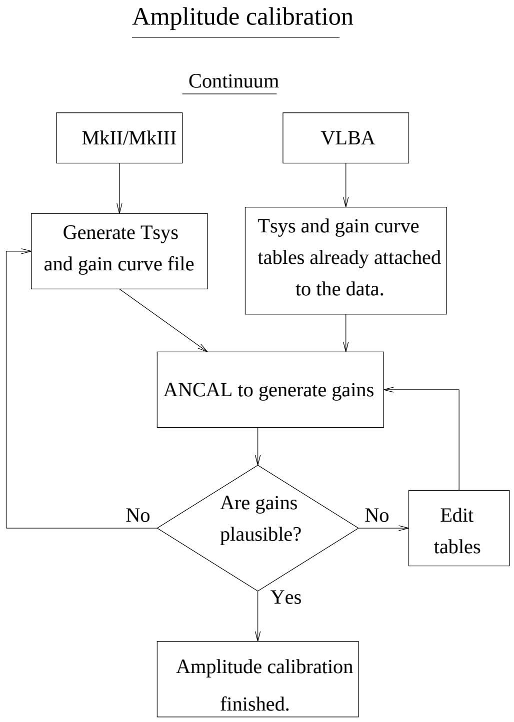

Figure 12.7: A block diagram of the procedure normally used to calibrate the amplitudes of continuum data in AIPS.

are fairly labour intensive on the part of the user. Figure 12.7 shows an overview of the amplitude calibration process for continuum data. For Mark III data a lot of effort has to be made on the part of the user to generate the ASCII file of system temperatures and gain curves. These are most often typed in by hand and frequently need to be modified and the calibration task re-run. Hopefully in the new era of the VLBA this aspect of amplitude calibration will soon disappear and the necessary calibration tables will be automatically attached to the dataset at the correlator.

Once the user is satisfied with the state of the ASCII file the AIPS task ANCAL is used to read the data in and generate the amplitude gain curves (see fig. 12.8). This curve can then be applied to the data in the standard fashion to produce visibility data that is then in units of Jansky.

When observing spectral line maser sources one can use the properties of the masers themselves to generate high quality amplitude calibration tables. As described in chapter 11 and shown in figure 12.9 one can use the total-power on the maser to track changes in antenna gain, atmospheric opacity, pointing etc. The AIPS task ACFIT uses the maser total power to generate the calibration table. Of course, one can also use the continuum method of calibration described

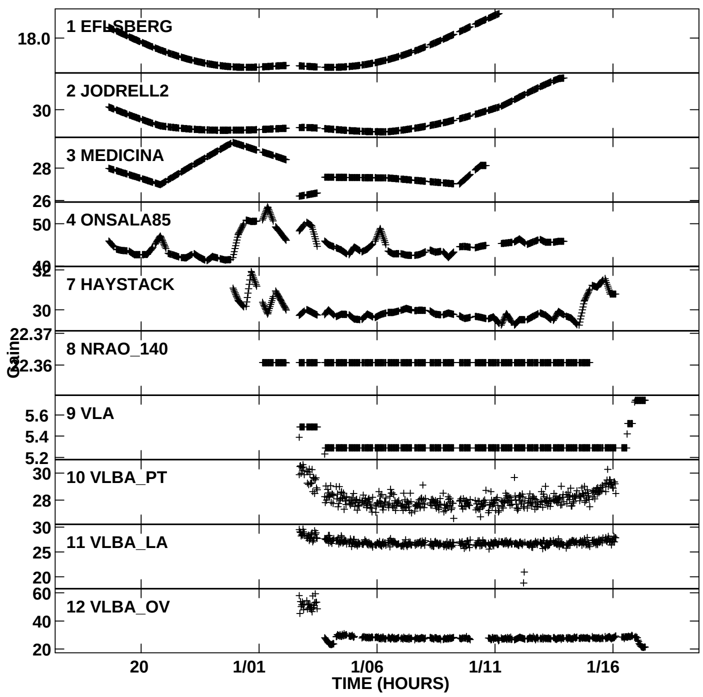

Figure 12.8: The calibration curves for the 10 useful antennas in the Cygnus A observation.

above to generate the same information but with a strong maser the spectral line method will produce more accurate calibration information. Some observing programs, especially those at 22 and 43 GHz, use short scans of strong masers at regular intervals to pin down the calibration for the continuum sources.

12.4.6 Fringe Fitting

Fringe-fitting is the method whereby the residual, unmodeled, delays and fringe rates in the data are determined and removed, thereby allowing users to average their data both in frequency and time. The theory of fringe-fitting is described extensively in chapter 10.

AIPS offers users two methods of fringe-fitting their data (see figure 12.10). The first, using the task FRING, utilizes a global fringe-search method that tends to be more sensitive than the traditional baseline-based method. The AIPS task BLING is a baseline-based fringe-fitting task that, under certain circumstances, has some advantages over FRING. The relative advantages and disadvantages of both methods are described in chapter 10.

The residual delay and rate results from either fringe-fitting task are stored in an AIPS *Solution* (SN) table. They can be edited or smoothed in a variety of

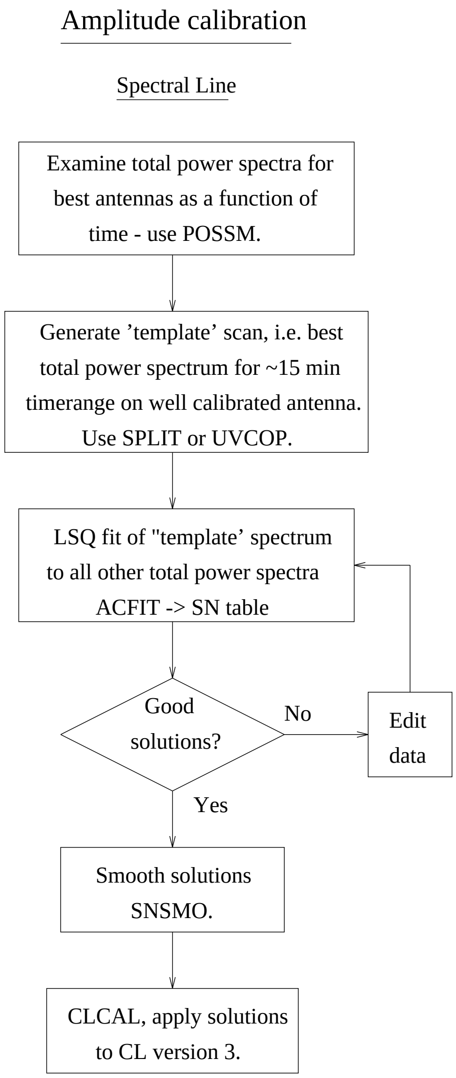

Figure 12.9: A block diagram of the procedure normally used to calibrate the amplitudes of spectral line data in AIPS.

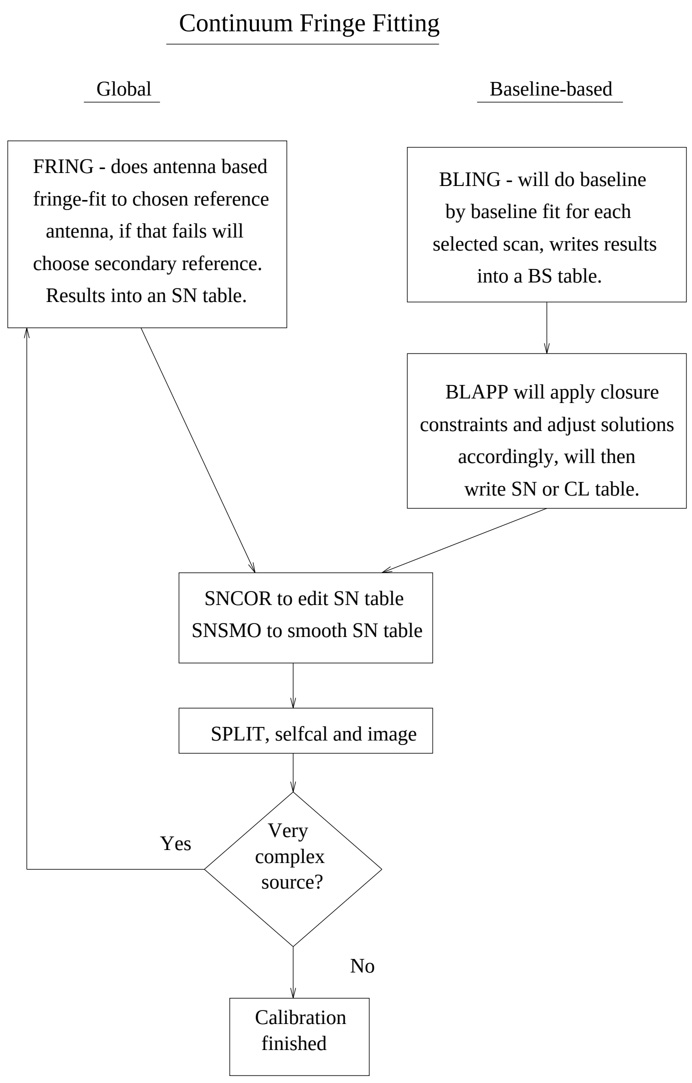

Figure 12.10: A block diagram of the steps involved in fringe-fitting a continuum observation.

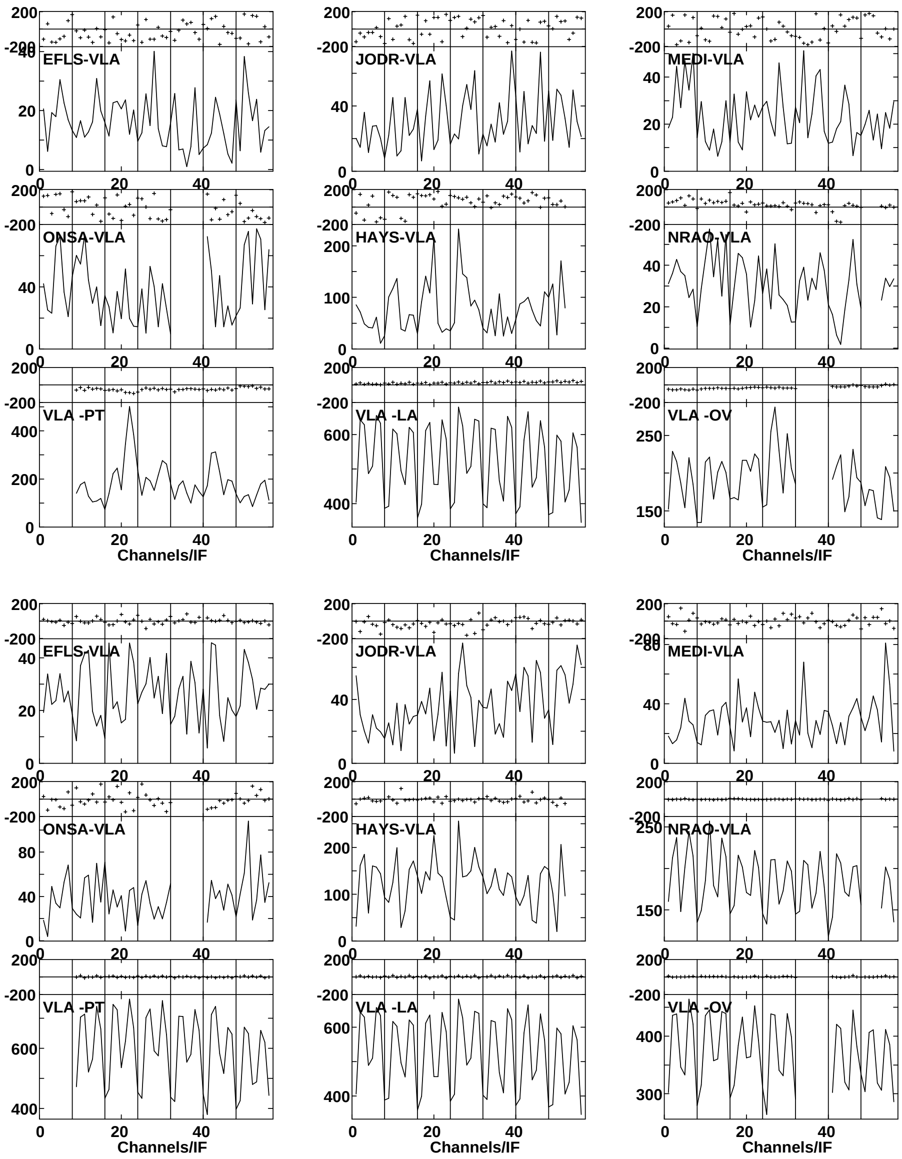

Figure 12.11: The upper plot shows the spectra on all baselines to the phased VLA for data on Cygnus A averaged for 5 minutes with all previous calibration (i.e. manual phase-cal and amplitude) applied. The lower plot shows the same data after a full multi-band fringe-fit was performed.

ways and then applied to the data in the normal fashion. Occasionally, the data may benefit from a second round of fringe-fitting. This would normally occur if the user was attempting to image a large and complex source in which the structure of the source itself contributed significantly to the residual delays and rates. In this case, after an image has been made, the data can be divided by the transform of the image and then re-fringe-fitted to produce more accurate values of the delay and rate errors. Figure 12.11 demonstrates how fringe-fitting corrects the data for the effect of residual delay and rate errors on the source Cygnus A.

Fringe-fitting can be a time consuming task and there are various shortcuts that a user learns by experience. Below is a list of hints that may or may not be useful to neophyte users of the AIPS global fringe-fitting task FRING:

- **REFANT:** try to choose a reference antenna that is present all (or most) of the time. Choose a big, sensitive, antenna if you have one or choose an antenna close to the geographical center of your array.
- **SOLINT:** the solution interval should be set to the coherence time (≈ 15 minutes generally). If it is $>$ the scan length FRING will adjust it to the scan length, if it is $<$ the scan length it will be adjusted to provide intervals of equal length.
- **APARM(5):** one of the crucial modes of FRING is controlled by this relatively obscure parameter. If APARM(5) = 0 then FRING will do a fit on each IF independently, if APARM(5) $>$ 1 FRING will perform a multi-band fit, if APARM(5) = 2 the multi-band fit will be performed and then the solutions will be adjusted for the differences in multi-band and single-band delays. If performing a multi-band fit, APARM(5) = 2 is recommended.
- **APARM(7):** the criteria for determining whether a FRING solution is real or not is based upon the signal-to-noise ratio. Set APARM(7) = 7.
- **delay-rate parameters:**
  - **DPARM(1):** the number of baseline combinations when "stacking" in the global search (see chapter 10). If you have a complicated source, i.e. a lot of structure, set DPARM(1) = 1, otherwise use the default value of 3.
  - **DPARM(2):** full width (ns) of the delay window to be searched. Users must be careful with this, if the data are weak and the window is set too wide, FRING can sometimes find a sidelobe in the delay resolution function and generate aliased solutions. The maximum allowed, and that used as the default, is $\frac{1}{\Delta\nu}$ where $\Delta\nu$ is the bandwidth of the individual spectral channels of each IF. Set DPARM(2) to -1 if you wish to switch off the delay search (often true for spectral line VLBI).

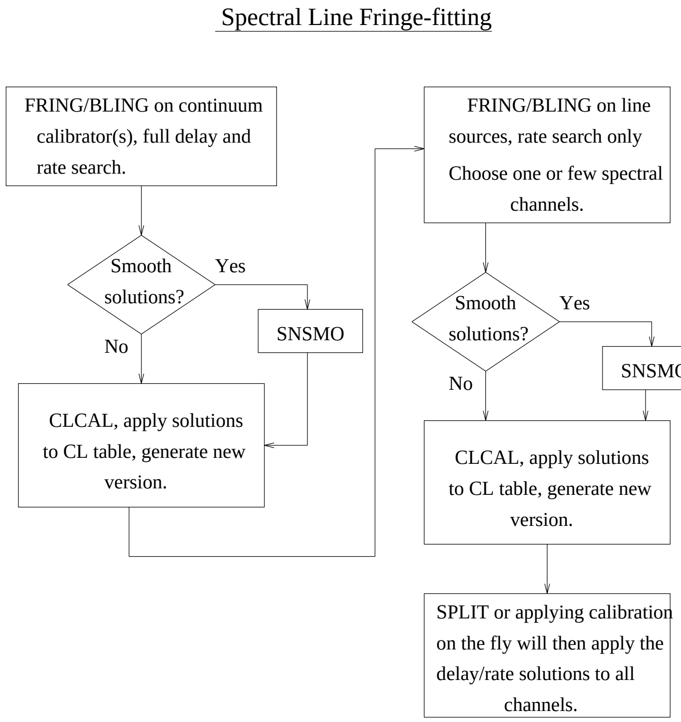

Figure 12.12: A block diagram of the steps involved in fringe-fitting a spectral line observation.

– **DPARM(3):** full width (mHz) of the fringe rate window to be searched. This should be kept as narrow as possible for signal-to-noise reasons. The maximum allowed is $\frac{1}{\tau_{int}}$ where τ_{int} is the integration time.

– **DPARM(4):** the data integration time τ_{int} (seconds).

The process of fringe-fitting spectral line sources is fundamentally the same as that for continuum sources. However, the details differ. Due to the nature of spectral line sources the residual delay cannot be estimated from the sources since the phase as a function of frequency is no longer a simple linear (or quadratic) function. What spectral line users do (see fig. 12.12) is to determine the delay error from continuum sources and then apply the correction to the spectral line source of interest. Then they choose a strong line (or group of lines) and determine the residual rate only on that line. This can then be applied to all other spectral channels allowing the data to be averaged in time.

12.4.7 The Final Product

Once a user is happy with the results of the basic VLBI calibration steps described above then the full calibration should be applied to the data in preparation for further self-calibration and imaging (see chapter 13). In AIPS this is done by running the task SPLIT and, for continuum data, usually averaging all spectral channels and IFs together to produce a data set with one channel per polarization. In the case of spectral line data spectral averaging is not usually done since the user normally wants to maximize the spectral resolution.

12.5 Conclusion

Hopefully, after reading this chapter, a user with little experience will have a basic idea of the steps to be followed in calibrating simple continuum and spectral line VLBI datasets. I have elected not to describe the steps required to calibrate polarization VLBI data because, at some level, this is still an evolving field. Similarly I have not covered phase-referencing VLBI. Both of these subjects are however covered in chapters 15 and 17.