Data Processing for the Australia Telescope

R.P.Norris

11 September 1987

This document reviews the off-line software which is yet to be written before we can produce maps from the AT. The main conclusion is that, although some mapping capability is available immediately, about 4 man-years of work remains to be done before we can do full justice to the high quality data from the AT. This implies that we will not be able to produce high quality maps until late 1989 with the present manpower.

1) INTRODUCTION

1.1) The scope of this document

In a few months time we may have our first astronomical data from the AT. This document reviews the necessary work remaining before we are in a position to process that data.

First, it should be pointed out that this document does not refer to our first tentative fringe searches on Day 1. For that task, it is proposed to use the PTILOOK software, which has been written for, and thoroughly tested by, the PTI and is well suited to the task. This document is instead concerned with how we will use the data once the AT (or some part of it) is running more or less routinely.

Second, it should be noted that, if calibrated uv data were available from the AT now, then the existing AIPS installation would in principle be able to produce maps from it. The existing installation would not, however, be able to calibrate the data, and it would be overwhelmed by some types of data that the AT is capable of producing. Specifically, mapping spectral line data with a reasonable number of channels would be an extremely painful process. Thus there are a number of tasks still ahead of us if we wish to process AT data effectively.

Third, this document considers only off-line software, although on-line software will be discussed briefly in the context of calibration in Section 6. A separate document discussing calibration in detail is in preparation.

It should also be noted that LBA data are not considered here explicitly. This is because the reduction of LBA data will not differ conceptually from that of CA data, so that little extra reduction software will be needed for the LBA. However, a great deal of processing software (i.e. to run the DPS and correlator) will be needed, which will represent an additional drain on manpower, detracting from that available for reduction software.

A number of new AIPS tasks are considered in this document, and are assigned provisional names. Note that some of these tasks are already being written elsewhere, and will not in general have the same names as those assigned here, so that many of the task names are liable to change.

1.2) Types of AT Data

The AT data to be processed can be divided into three broad categories: synthesis data, non-synthesis astronomical data, and primary calibration data. Non-synthesis astronomical data includes astrometric data and data from a variety of experiments using only one or a small number of baselines. Primary calibration data in this context consists only of data taken in order to determine the parameters of the instrument . It does not include those secondary calibration scans that will normally be inserted periodically throughout a synthesis observation. The issue of calibration will be considered in outline in Section 6 of this document.

1.3) Philosophy

It is intended that all the off-line data processing should be done within the AIPS environment. The philosophy behind the suggested scheme rests on the following assertions:

a) Any calibration of the data should always be reversible, so that a second, better calibration can be applied subsequently if available.

b) Most data should not need re-calibrating, so that the end-user (the astronomer) need not worry about calibration. Nevertheless, this does not preclude him from applying some favourite calibration scheme in order to improve his maps.

c) The end-user should not <u>need</u> to know about the instrumental details. Note, however, that this does not preclude him from utilising any such knowledge he does have in order to improve his maps.

d) The calibrated data should be transportable to any other AIPS site, and should not need any AT-specific tasks to process it.

1.4) Layout of this document

This document is arranged as follows. Section 2 details the hardware that is expected to be available, Section 3 describes the data transfer mechanism, Sections 4 and 5 consider the data paths for each of the types of astronomical observation listed above, and Section 6 considers the issue of calibration. Finally, Section 7 lists the tasks which are still lay in front of us before we have a working system, considers the manpower requirements, and recommends a scheme of implementation.

2) HARDWARE

2.1) On-line processing

Figure 1 shows the hardware that is expected to be available to the AT by 1988. At Culgoora, the data from the antennas is processed by the correlator and its associated array processor (AP), and the resulting data are written to a transport medium which may ultimately be an optical disk but which is likely in the early days to be a magnetic tape.

The processing by the correlator / AP includes calibration of the data, using the best set of parameters available at the time of observation. These calibration parameters will be held in a calibration database residing on the Culgoora SYNCH (VAX 8250) computer. A copy of this database will also reside on the VAX 750 at Epping.

2.2) Off-line processing

Off-line processing will generally be done on the main Data Reduction Computer at Epping. However, it may also be done on the off-line VAX (ASYNCH) at Culgoora, which will have a copy of AIPS maintained on it. Alternatively, the data may be transferred to a conventional FITS tape to be transferred to another site. This document concentrates on the route through the main Data Reduction Computer at Epping.

At the time of writing the choice of the Data Reduction computer had not been made. As a result, development of reduction software now would be inefficient, and has therefore largely been shelved pending a decision. However, it is assumed throughout this document for the purposes of generality that the machine will be a Unix machine (e.g. Alliant, Convex, Gould, etc.). This machine will be referred to as the DRC (Data Reduction Computer). Similarly, it is not known what imaging facilities will be available, although it is probable that a small number of workstations (in the Sun/Apollo/Mac II class) will be available. The DRC, together with its displays, workstations, and peripherals, is intended to be able to handle all the off-line data processing and imaging.

2.3) The Cyber 205

It will be noted that the Cyber 205 is absent from this scheme. This is because (a) it is not clear, given our probable purchase of a mini-supercomputer, that the power of the 205 will be required in the immediate future, and (b) experience has shown that the 205 installed at CSIRONET is not yet suitable for routine data processing. In particular, bugs in the compilers and operating systems change on a time scale of weeks, so that a working program does not remain working for very long without extensive maintenance. This situation may change in the future, in which case this decision may be reviewed.

3) DATA TRANSFER

3.1) Hardware

The data transport medium from Culgoora to Epping will probably ultimately be an optical disk. However, this is not likely to be cost-effective in the early days of AT operation, and so it is expected that mag tapes will be used initially, and a transfer to optical disks made at some later date. The data transfer format, RPFITS, is suited to any such medium.

3.2) Software

A draft of the Data Transfer format, RPFITS, was presented in an earlier AT document (AT/25.1.1/028). Since then, there have been a few modifications to the format although the major design features have been retained. RPFITS is based on the FITS format (Greisen and Harten, 1981, Astr. Astrophys. Suppl., **44**, 371) , and differs primarily in that (a) many scans can be included in one file, (b) the block size of the file is changed to give a higher speed when transferring data to and from peripherals, (c) it is more robust against hardware failures, and (d) it uses IEEE REAL*4 format.

A major modification to RPFITS, involving the use of extension tables, has been occasioned by the desire to be able to use existing and forthcoming AIPS tasks as much as possible. In the original RPFITS specification it was stated that extension tables would not be permitted. It has now been decided that extension tables containing the calibration data in the form of gain tables should accompany the data. Similarly, any other tables may be appended. However, in order that the data be robust against computer crashes, the following scheme is proposed.

During an observation, calibration data will be read from the database by SYNCH, and passed on to the correlator for calibrating the data. The correlator will then calibrate the data and write it in RPFITS on to the recording medium. Simultaneously, SYNCH will write an entry in gain tables stored as VMS files on magnetic disk. These entries will be written at an interval shorter than the timescale of changes in the calibration, which may typically be once per minute. At the end of each scan, SYNCH will pass each table to the correlator which will then write it onto the recording medium after the data from that scan. In the event of a failure of the recording medium, the tables may be regenerated from the calibration database.

4) SYNTHESIS DATA

This category of data is likely to form the bulk of the data processing load. AIPS already contains most of the tasks required for processing this data, except that in many cases the existing tasks are limited in their capability to handle spectral line data. These limitations are currently being examined, and will hopefully be rectified, by the AIPS team at NRAO.

The data path for astronomical observations is shown in Fig. 2, and a detailed data processing route for synthesis data within the DRC is shown in Figure 3. Most of the tasks shown in Figure 3 already exist. I now consider each of them in turn.

4.1) RPLOD

This task already exists but needs some modification (2 man-weeks) to cope with the modified RPFITS format.

4.2) RECAL

This task will be able to take calibrated data from the AT and re-calibrate it using a selected calibration set from the calibration database. To do so, an associated task (CALREAD) must be written to access the database (in the VAX) and transfer the selected data (via ETHERNET) to the DRC. This function is considered in Section 6.

4.3) POSSUM

This task produces a spectrum from the raw u-v data. It thus enables an astronomer to decide which half-dozen of his 8192 spectral channels actually contain data and are worth mapping. This program is currently being written at NRAO.

4.4) EDIT

We will require an interactive data editor, with the following functions. Data can be displayed in a variety of ways on an interactive device, and data can be selected with a mouse or trackball. Selected data can then be edited or displayed at an enlarged scale. The task should also contain a selection of automatic editing features (e.g. clipping). It is believed that such a task is already being written at NRAO.

4.5) UVSRT, UVMAP, CLEAN, MX, VM

All these tasks already exist within AIPS. However, most of them do not cope adequately with spectral line data, and so need modification. It is understood that this is already being attacked at NRAO. No doubt there will be other changes required for the AT (ASCAL, for example, contains some VLA-specific code). We should allocate perhaps 6 man-weeks to modification of these tasks.

4.6) Others

There are a number of tasks which would be desirable, although not essential, for making best use of AT data.

4.6.1) BWSMAP

This task will use the extra information provided by the wide-band observations of the AT to fill in some of the missing uv data. It thus effectively uses bandwidth synthesis techniques and fits to a spectral index map as well as an intensity map. Such a task will be invaluable to the AT. Fortunately, such a task is already being developed at Jodrell Bank.

4.6.2) 3DCLEAN

This task is similar to BWSMAP in that it considers the frequency axis as well as the spatial axes in the uv data. However, it differs in that it does not assume a power law spectrum in the frequency domain, but instead attempts to CLEAN in the frequency domain at the same time as CLEANing in the spatial domain. The effect of this is to remove artefacts frequently seen in VLA spectral-line maps caused by ringing in the frequency domain, but at the same time maintain (and even increase) the spectral resolution. This task is not yet being developed anywhere, and will require considerable manpower (perhaps 1 man-year). Thus, in the immediate future it will remain on our wish-list.

4.6.3) SDFLUX

This task will add the single-dish fluxes obtained at Parkes to the AT data. This should not require more than 2 man-weeks to implement.

4.6.4) ATMAP

The present UVSRT/UVMAP gridding combination can be improved in two ways. First, when processing data on a mini-supercomputer, most maps (i.e. all but the very largest) can be stored entirely within core, so dispensing with the need for UVSRT before gridding. Second, gridding in UVMAP can be speeded considerably (and UVSRT dispensed with) if we take advantage of our linear compact array, and grid radially rather than scanning. To write a task (ATMAP) to cover both these cases will require perhaps 6 man-weeks.

4.7) Output Display

AIPS at present supports a wide variety of graphics and TV displays, which are adequate for our needs. However, it is likely that there will be additional tasks necessary (estimated 6 man-weeks) to make full use of the workstations which we will need to display images form the DRC. This assumes that there already exist the appropriate Y-routines.

5) NON-SYNTHESIS ASTRONOMICAL DATA

The format of this data will be similar to that produced by the Parkes–Tidbinbilla Interferometer (PTI), and a suite of software (PTILOOK) already exists to process this data. However, it is planned to incorporate all the PTILOOK software into AIPS, after which the data route will be that shown in Fig. 2. The data produced by the AT will fall into the following categories:

5.1) Fringe Searches

This type of observation simply measures how much flux is being emitted from compact components at a particular position in the sky, and is the mode most often used on the PTI. For example, all the observations of SN1987A are of this type. In general, it may be used to survey a large number of sources (e.g. IRAS galaxies, flat-spectrum quasars) either to gain astrophysical statistics, or else to see which of them have enough flux to be mapped by the AT. The data processing paths for this type of observation are shown in Fig 4. The only new task, SEARCH, is a task to transform uv data to fringe frequency/delay space. This task is already implemented in PTILOOK, and needs only to be transferred to AIPS, requiring about 2 man-weeks.

5.2) Astrometry

This category of data is already being produced by the PTI, and will continue to be produced throughout the lifetime of the AT both on LBA and CA baselines as (i) part of the AT calibration , requiring (ideally) unresolved sources at known positions with known fluxes, and (ii) for relating radio results to results in other wavebands. The data processing route is illustrated in Fig. 5. At present we are still experimenting with analysis techniques on the PTI data, but will probably use the technique of fitting the phase equation to the absolute phase of a calibrator. At present, because the PTI local oscillators are not phase-stable, this technique has to be preceded by a stage of subtracting off the phase of a nearby calibrator source with known position. This latter technique may continue to be needed for LBA data, and so will be retained in the AIPS implementation.

An additional technique, that of using group delay, has not been used yet on the PTI because of the small bandwidth available. However, this may prove to be a valuable technique on the wide bandwidths available on the AT. We therefore need to consider software for analysing this type of astrometry data. The required tasks are therefore as follows.

5.2.1) PHREF

This task will subtract off the phase of a reference source from that of a target source, and to do so it will in general have to interpolate between the observations of the reference source. The referencing itself will be achieved by modifying the gain tables, as is already done in AIPS by the task ASCAL. However experience with the PTI has shown that a splined interpolation is considerably more successful that the simple linear interpolation done by ASCAL, and so PHREF will allow some flexibility in the interpolation. This task already exists in PTILOOK, but will need appreciable modification to run in AIPS. It will also be written so as to incorporate the spectral-line phase-referencing requirements. It is possible that an interactive task (using the TV and cursor) would be even more successful than a spline interpolation, so that a considerable time (e.g. 4 man-weeks) must be allocated to this task.

5.2.2) SINEFIT

SINEFIT will be a task that not only fits a sinusoid (the phase equation) to phase reference data, but will also remove lobe ambiguities that can mislead such fitting. This task already exists in PTILOOK, and it has been found necessary to make it an interactive task. Thus transferring it to AIPS will require use of the TV and cursor, and will require about 4 man-weeks.

5.2.3) GRPFIT

GRPFIT will be a task that will measure the group delay on a source by looking at the phase as a function of frequency. It will be useful for astrometry using wide-band (or multi-frequency) AT data. This major task is already encoded in various VLBI packages, and needs transferring to AIPS. It is estimated that this will take about 4 man-weeks.

5.3) Spectral-line maser spot mapping

Maser sources can be mapped effectively with only one baseline, by assuming that the masers are unresolved point sources, and then measuring their positions relative to a reference maser feature. Of course, synthesis arrays can do this even better, but single-baseline observations have the advantage of requiring less telescope (and computer) time, so that a larger number of sources may be mapped with a given allocation of resources. Fig. 6 shows the data processing path for this mode of observation, and it can be seen to be very similar to that for astrometry. In fact, by making the astrometry tasks described above suitably general, no extra tasks need be written for this mode of observation.

5.4) Other observing modes

In addition to the modes described above, non-synthesis observations may be used for a number of other experiments such as measuring pulsar proper motions, modelling single-baseline data in terms of simple spatial structures (e.g. surveys for jets), and geodesy. However, each of these functions is so specialised that it does not seem appropriate for us to include them in the critical path of AT data processing. Instead, software for these functions can be relegated to a position further downstream.

6) CALIBRATION

6.1) The calibration process

There are two parts to the calibration of AT data. First, entries must be placed in the calibration database (used here in a general sense of including both Rdb and non-Rdb data). Some of these entries are derived from astronomical measurements (e.g. baseline offsets), some from the literature (e.g. ephemeris information), and some are from control/monitor information (e.g. meteorological conditions, T_{sys}). Secondly, this calibration information must be extracted from the database and applied to the data.

The resulting process is illustrated by Fig. 7. The main database is maintained at Culgoora on SYNCH, and is updated automatically by (a) on-line software (not considered here) which enters data from control/monitor points, (b) manual entering of data from the literature, (c) from primary calibration observations (considered in Section 6.4 below), which might be analysed either at Culgoora or at Epping, and (d) from secondary calibration observations (considered in Section 6.5) below.

Extraction of the calibration data, and correction of the astronomical data, can occur in one of two ways. First, the data are always corrected by the on-line software in the correlator / AP. This is discussed in AT/25.1.1/002 and is not considered here. Second, the data may be re-calibrated either at Culgoora (in ASYNCH) or at Epping, using the RECAL task, which has already been described in Section 4.2.

Calibration of data within AIPS is presently done using gain tables. To ensure maximum compatibility of our tasks with those written by NRAO, all calibration tasks will continue using gain tables. By taking this approach, rather than allowing AIPS tasks to access the database, we can also ensure that both AIPS tasks and database tasks can be worked on separately in a stand-alone mode, thus maximising ease of writing and debugging these tasks. We thus need separate tasks to access the database from AIPS. These are called CALWRITE (to write the contents of a gain table into the database) and CALREAD (to read a selected portion of the database into a gain table). It is expected that each of these will take about 4 man-weeks to write, bearing in mind that both involve accessing the VAX database from the DRC.

6.2) The Calibration Database

The calibration of AT data is based on a body of data held in the Calibration Database. This database is stored using the Rdb package on the Culgoora SYNCH VAX 8250. A secondary copy will also be kept on the VAX 750 at Epping. The database contains a number of distinct categories of calibration data, as follows:

1) Geometrical data - This includes the location and orientation of antennas, stations, etc. This is normally updated only at a configuration change, but can in principle change at any time.

2) Instrumental data - This includes the complex gains, cross-polarisation, T_{sys}, etc. of the various receivers and other bits of hardware, as determined by the one-bit correlators and phase-switched detectors included specifically for this purpose. This is normally updated whenever a calibrator is observed, as it may change on a timescale of hours.

3) Ephemeris and geophysical data - This includes all the data obtained from the ephemeris, USNO circulars, etc., and is updated periodically by reference to the literature.

4) Meteorological data - This includes data from the local meteorological instruments, and possibly ionosonde data, and is updated on a time scale of minutes-hours. It is not yet known whether Rdb will cope with the volume of meteorological data. At present it is assumed that it will, but that these data may need to be purged at (perhaps daily) intervals.

5) Empirical data - This includes all effects not covered above. This includes random atmospheric and ionospheric effects.

6.3) Calibration Techniques

The first step of calibration is the process of placing entries in the calibration database (which includes both the Rdb database and the gain tables stored by SYNCH). We may distinguish three types of calibration:

(a) Non-astronomical calibration, which includes all calibration data not obtained from astronomical measurements, and includes not only categories (3) and (4) above but also category (2).

(b) Primary calibration, which is the determination of geometrical parameters (category (1) above) by observing calibrator sources. In principle this includes pointing solutions for individual antennas, but in practice pointing is probably best regarded as a separate process. Primary calibration is discussed in Section 6.4 below.

(c) Secondary calibration, which is the determination of the residual complex gains of the receivers after compensation has been applied for all other known effects (category (5) above). This calibration is achieved by interspersing observations of calibrator sources at (typically 10-minute) intervals throughout a synthesis observation, as discussed in Section 6.5 below.

6.4) Primary Calibration

The data path for primary calibration data is shown in Figure 8, and the software path inside AIPS is shown in Figure 9. The function of this data is to calibrate the geometrical parameters of the array. It is believed that the most efficient way of calibrating these is to observe a large number of calibrator sources widely spaced in the sky, and then solve simultaneously for all the free parameters. In some cases, some of the parameters will have been determined by other means, and so this task, designated CALFIT, must also allow any of the parameters to be designated free or fixed. Because of the large number of parameters, they are not well suited to AIPS adverbs, and so the input and output of CALFIT should be in the form of ASCII files and gain tables.

Such an approach is already provided by the JPL program MASTERFIT, but we will need to write our own version, tuned to the requirements of the AT, almost from scratch. To write such a task will be a major effort, involving considerable experimentation and requiring, at a guess, about 6 man-months. In addition, the task CALWRITE, as discussed in Section 6.1, will be needed to transfer the output of CALFIT into the calibration database.

6.5) Secondary Calibration

This section is concerned with the secondary calibration data regularly interspersed between scans of synthesis data. Secondary calibration aims to correct synthesis data for effects which vary on a timescale of hours or less, as opposed to the primary calibration parameters which remain constant on this timescale. Secondary calibration differs further from primary calibration in that no cause is necessarily assigned to a phase error, but that instead all corrections are purely empirical. Thus secondary calibration is achieved by observing a calibrator source as close as possible to the target source, so that they share common atmospheric and ionospheric delay errors. The calibrator source is observed typically every few minutes, the complex gain of each antenna determined from these observations, and the resulting complex gain errors interpolated over the observations of the target source and written into the database. Any residual errors in the primary calibration is also removed by this technique.

Since only on-line software is required for this process, this task does not really fall within the scope of this document, and is included only so that a complete picture of the calibration process can be given. It will be considered in more detail in a separate document in preparation.

7) IMPLEMENTATION

As well as the tasks considered above, considerable effort will be needed for the following tasks:

a) Install AIPS on the DRC and associated workstations (6 man-months, including writing new Q, Y, and Z routines)

b) Write tasks to maintain and access the calibration database (3 man-months)

Table 1 shows the total manpower required for each task. The total manpower required to be able to process AT data (excluding priority 2 options) on the DRC amounts to about 100 weeks, where one man-week means a full 40 hours spent solely on that task. Because of other duties, no person in the present AT software group is able to spend more than about half that time on writing software. Thus a realistic estimate might be that 4 man-years are required, where a man-year in this case means one person spending a substantial fraction of his time for a year on writing software. Given the three people available at present, it appears that we will not be able to process AT data until late 1989. In addition, it is likely that the estimated times are under-estimated because of the additional overheads required in famailiarising ourselves with the new DRC and its operating system.

In order that the data reduction facilities should be available in late 1988, when the first astronomical data will be emerging from the AT, it is necessary to recruit additional manpower. At present, all suitable people within the Division are already heavily committed to writing on-line software, and so an external appointment is necessary. If we could recruit someone by the end of 1987, then the data reduction software could be ready by early 1989. Before then, it is proposed that the current work on the database and the PTI should continue, so that when we purchase the DRC at the end of this year we will be in a position to start implementing the tasks described herein.

Task	Priority	Manpower(weeks)
RPFITS format	1	2
RPLOD	1	2
UVSRT, etc.	1	6
SDFLUX	2	2
ATMAP	2	6
SEARCH	1	2
PHREF	1	4
SINEFIT	1	4
GRPFIT	2	4
CALFIT	1	26
CALWRITE	1	4
CALREAD	1	4
Install AIPS on DRC	1	26
Database maint./comms.	1	13
Graphics for workstations	1	6
TOTAL	**1**	**99**
TOTAL	**2**	**12**

Note: Priority 1 tasks are required in order to calibrate AT data and produce maps. Priority 2 tasks are not essential but should be implemented as soon as possible after the Priority 1 tasks are completed.

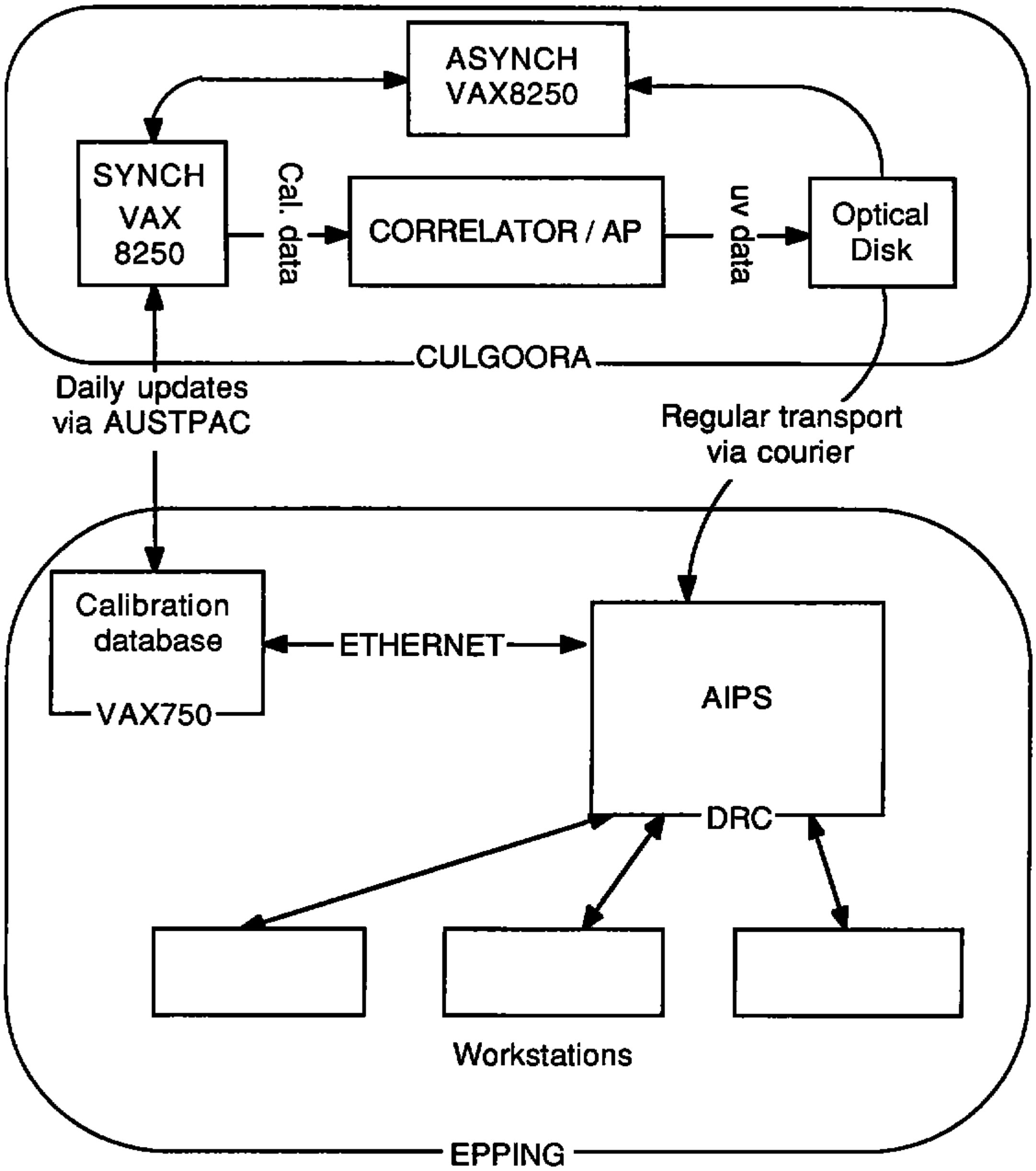

Figure 1 : AT Data Reduction Hardware

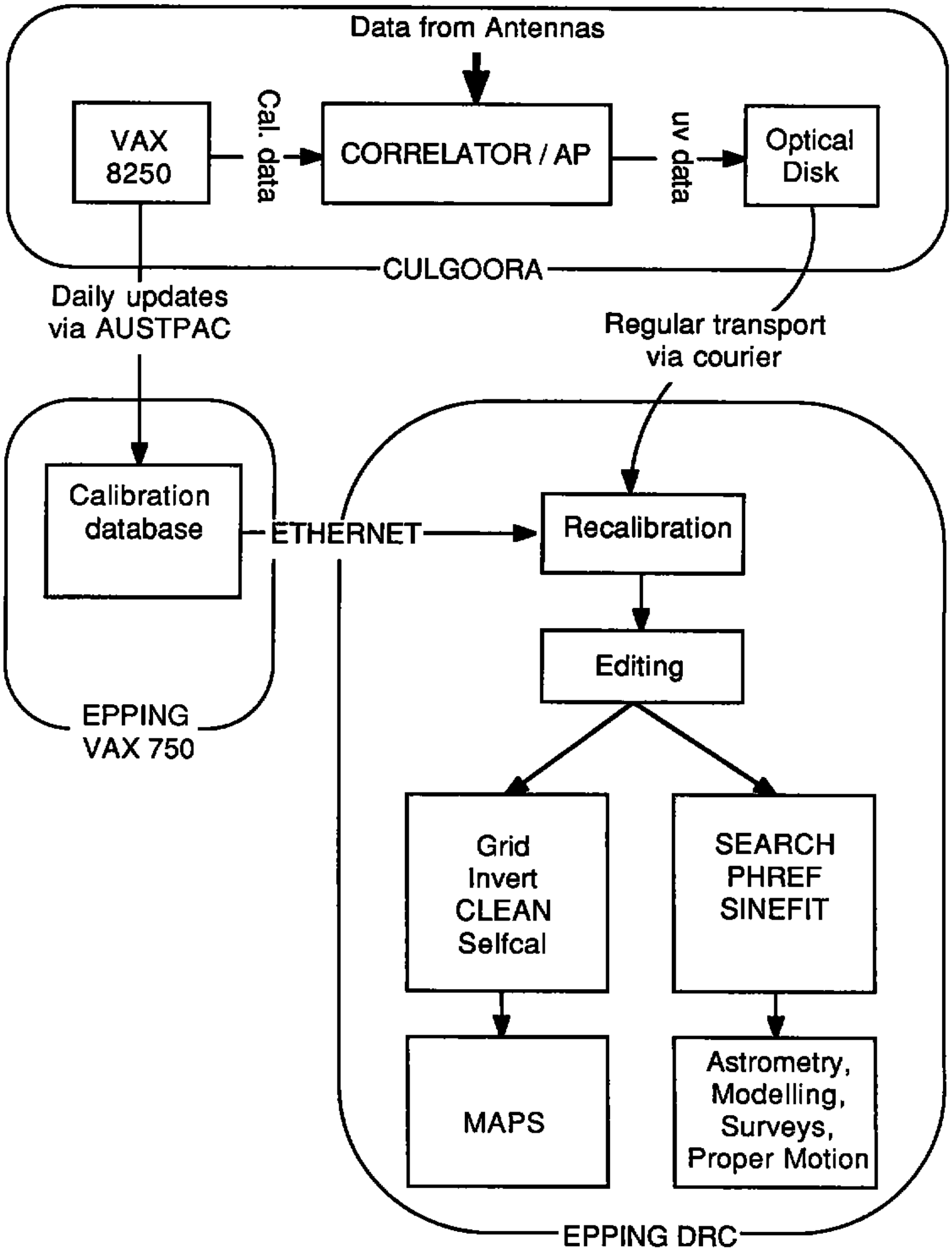

Figure 2 : Data paths for Astronomical Data

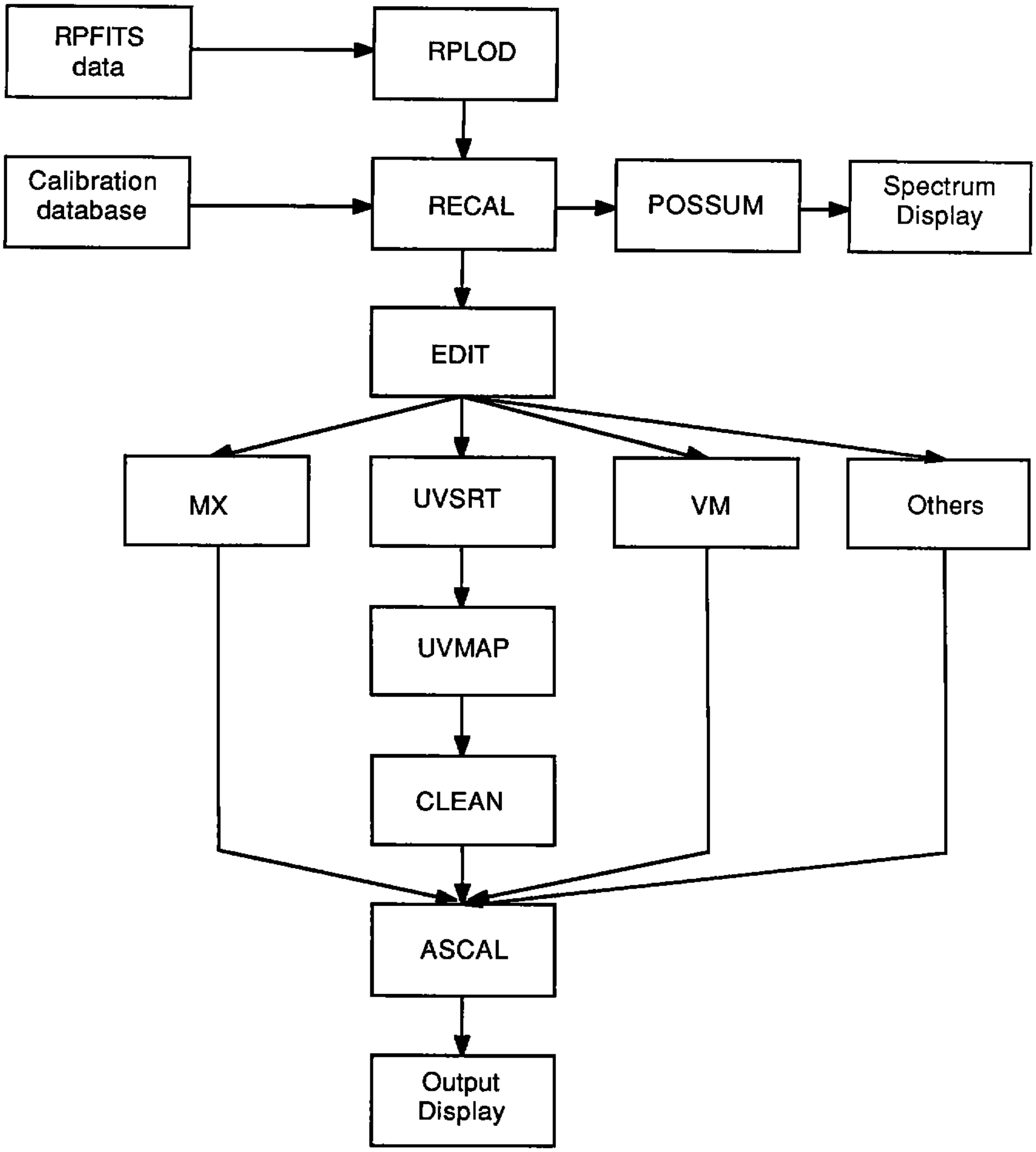

Figure 3: Synthesis Data Processing Path

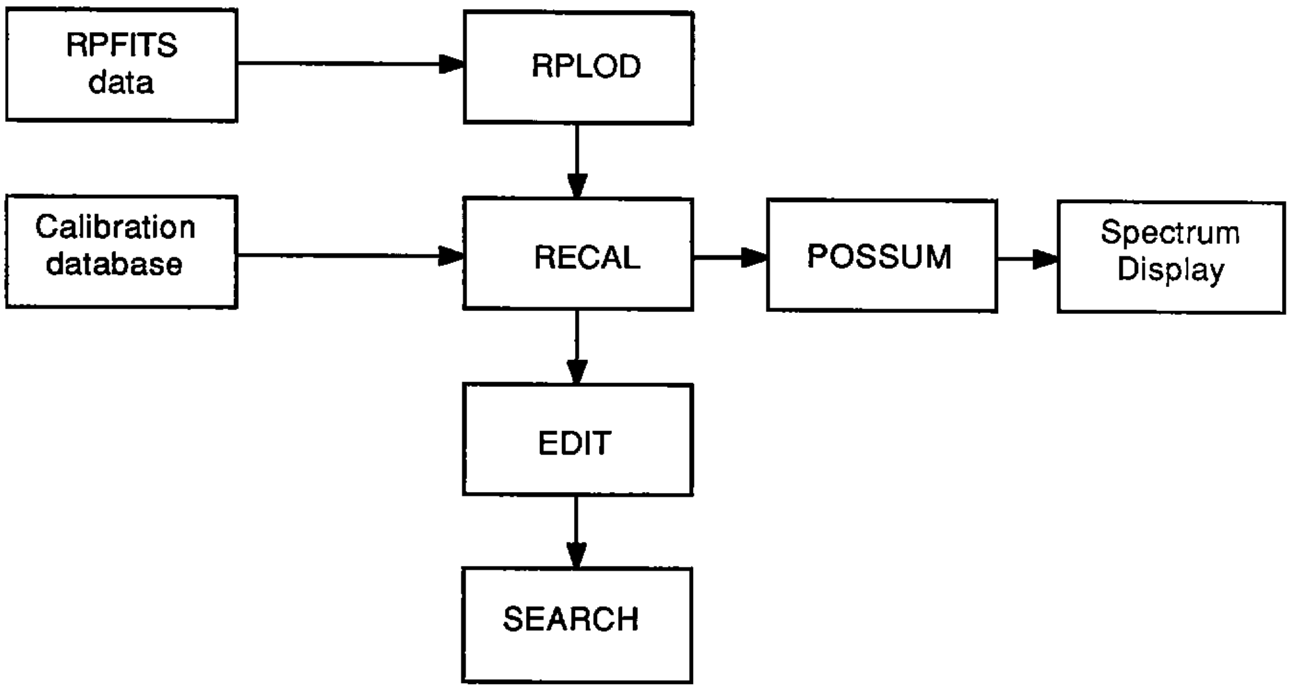

Figure 4: Fringe Search Data Processing Path

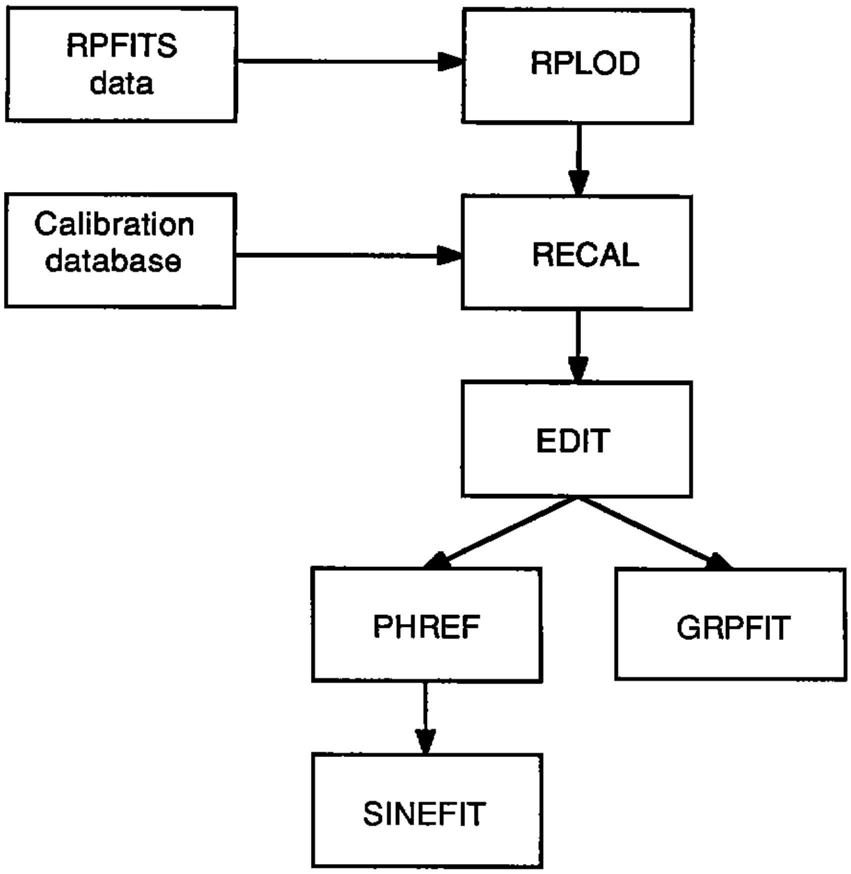

Figure 5: Astrometry Data Processing Path

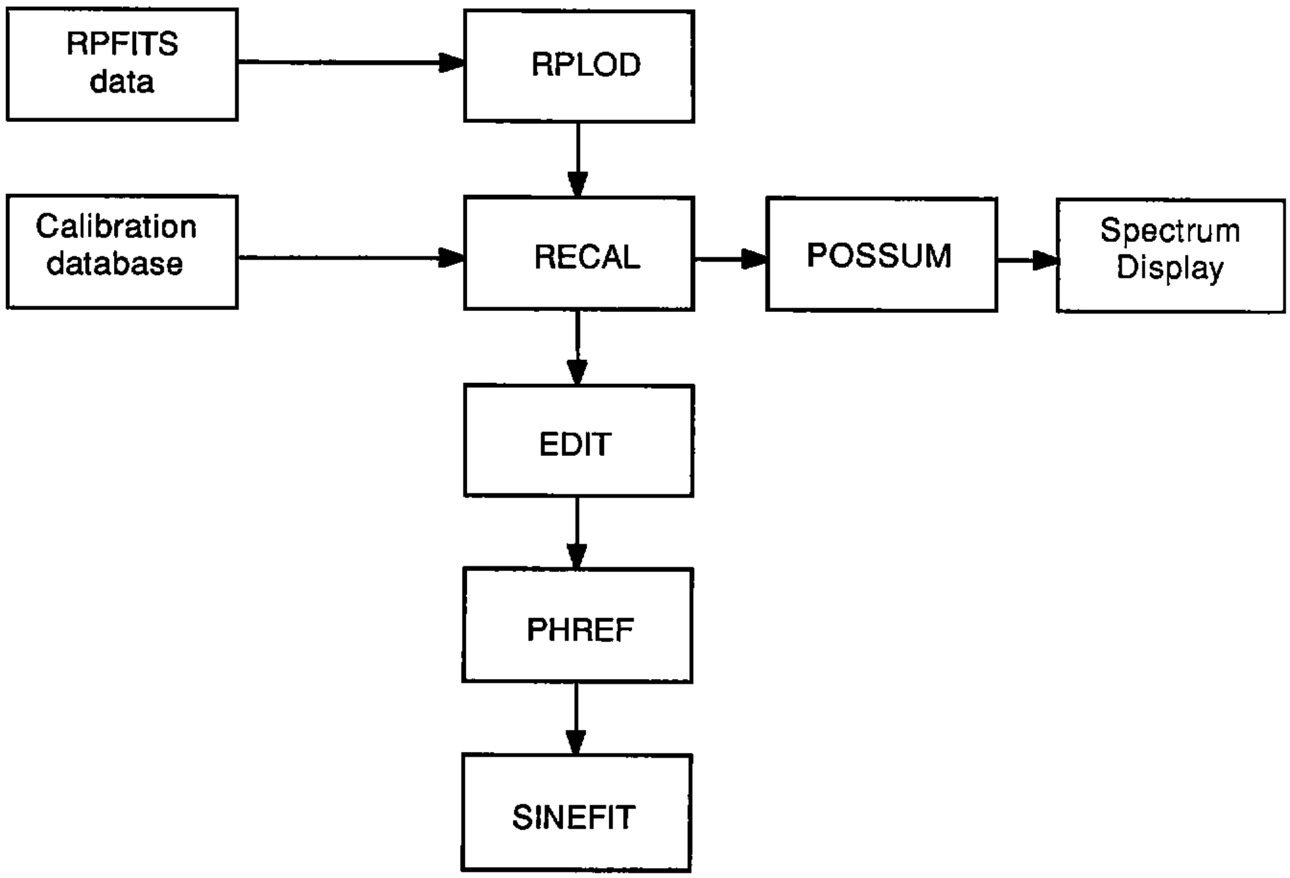

Figure 6: Maser Spot Map Data Processing Path

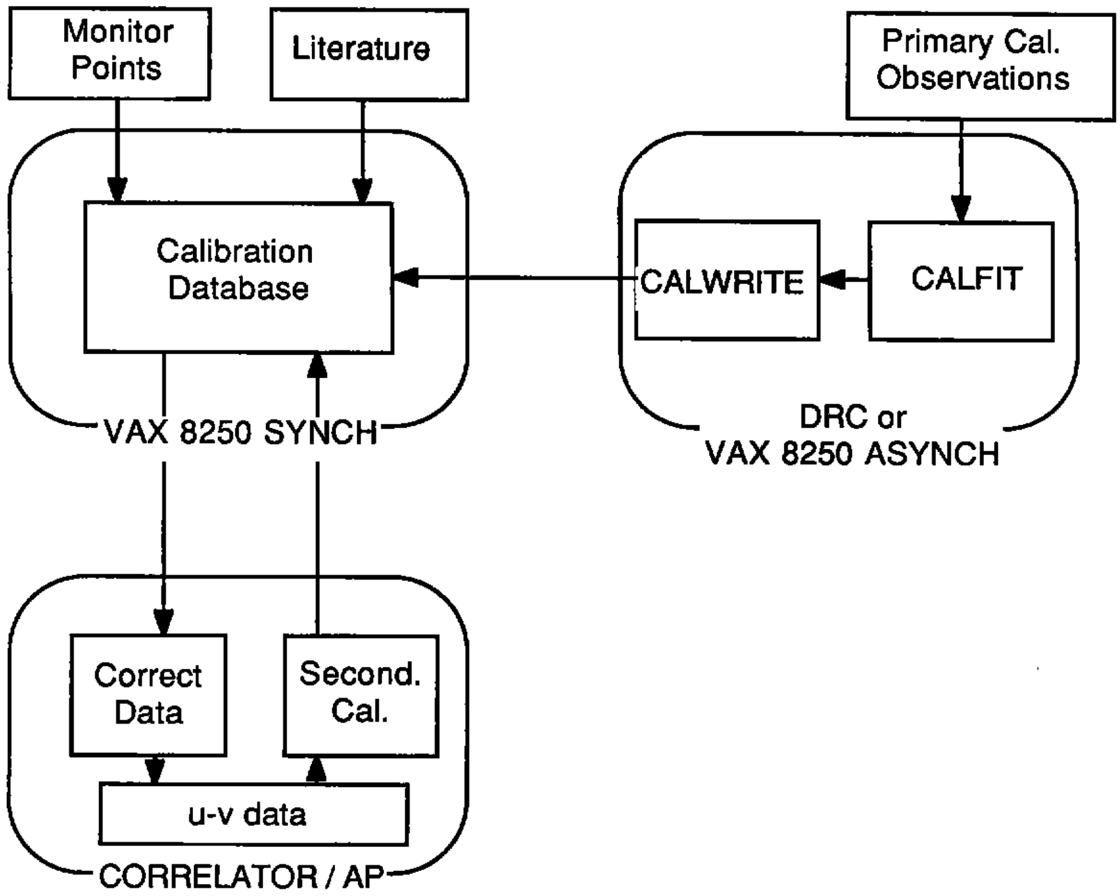

Figure 7: Data Calibration

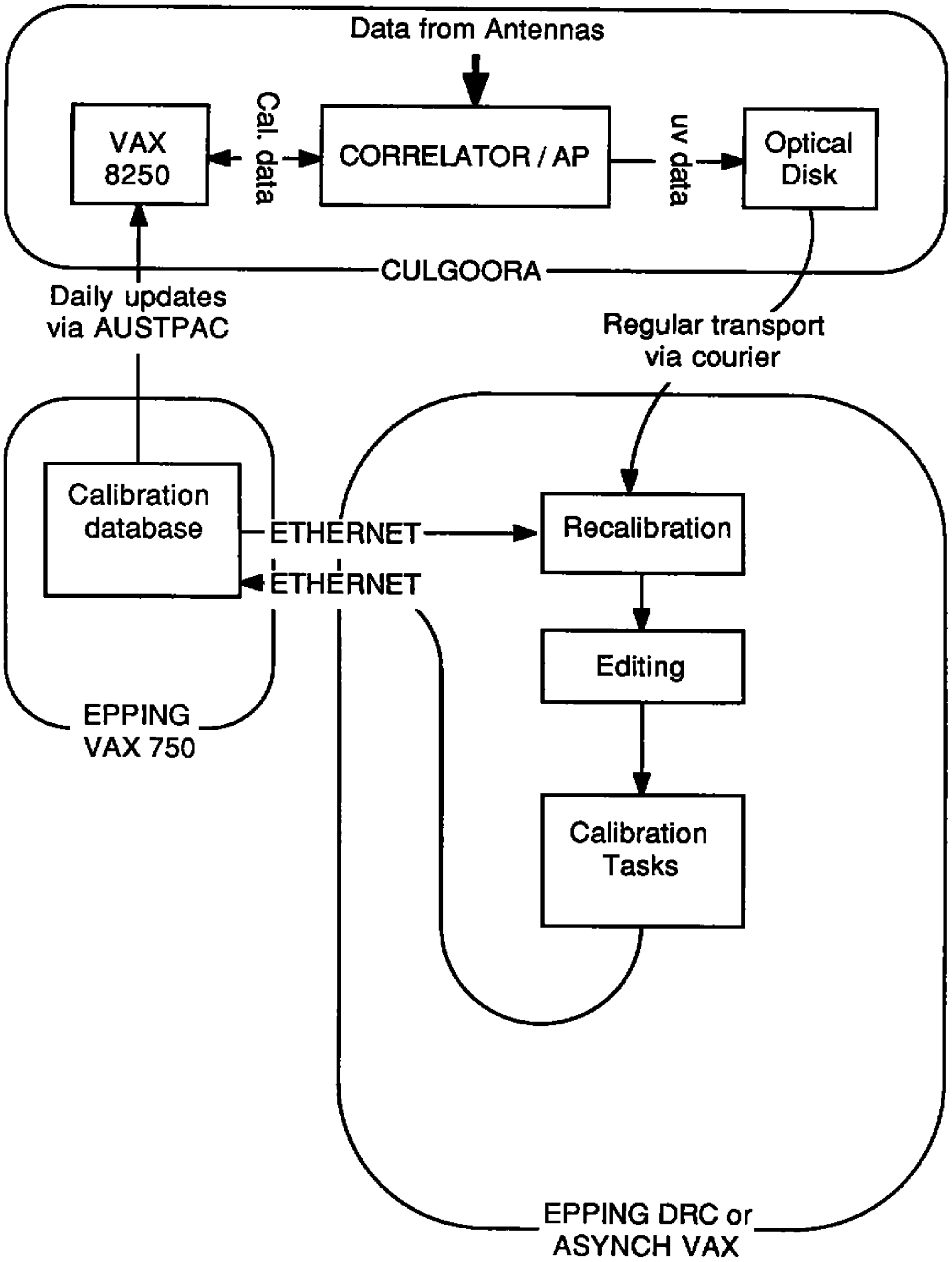

Figure 8 : Data paths for Calibration Data

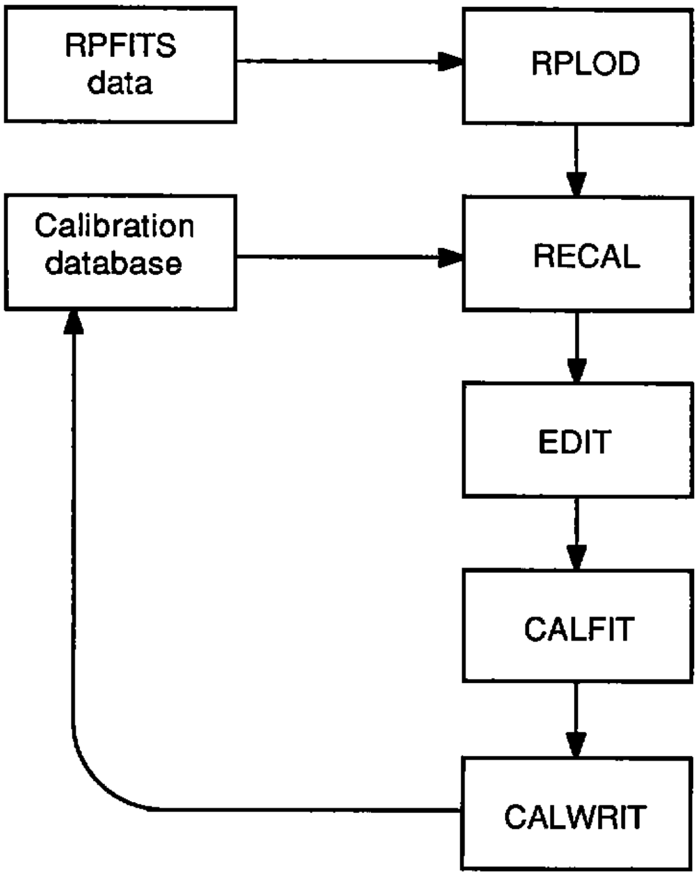

Figure 9: Primary Calibration Data Processing Path