VLB ARRAY MEMO No. 125

ASTROMETRIC DATA PROCESSING WITH A VLBA

David Shaffer
Phoenix Corporation

This memo has been written for the VLBA data processing group, but it is being sent to several other groups in order to raise their consciousness about astrometric requirements.

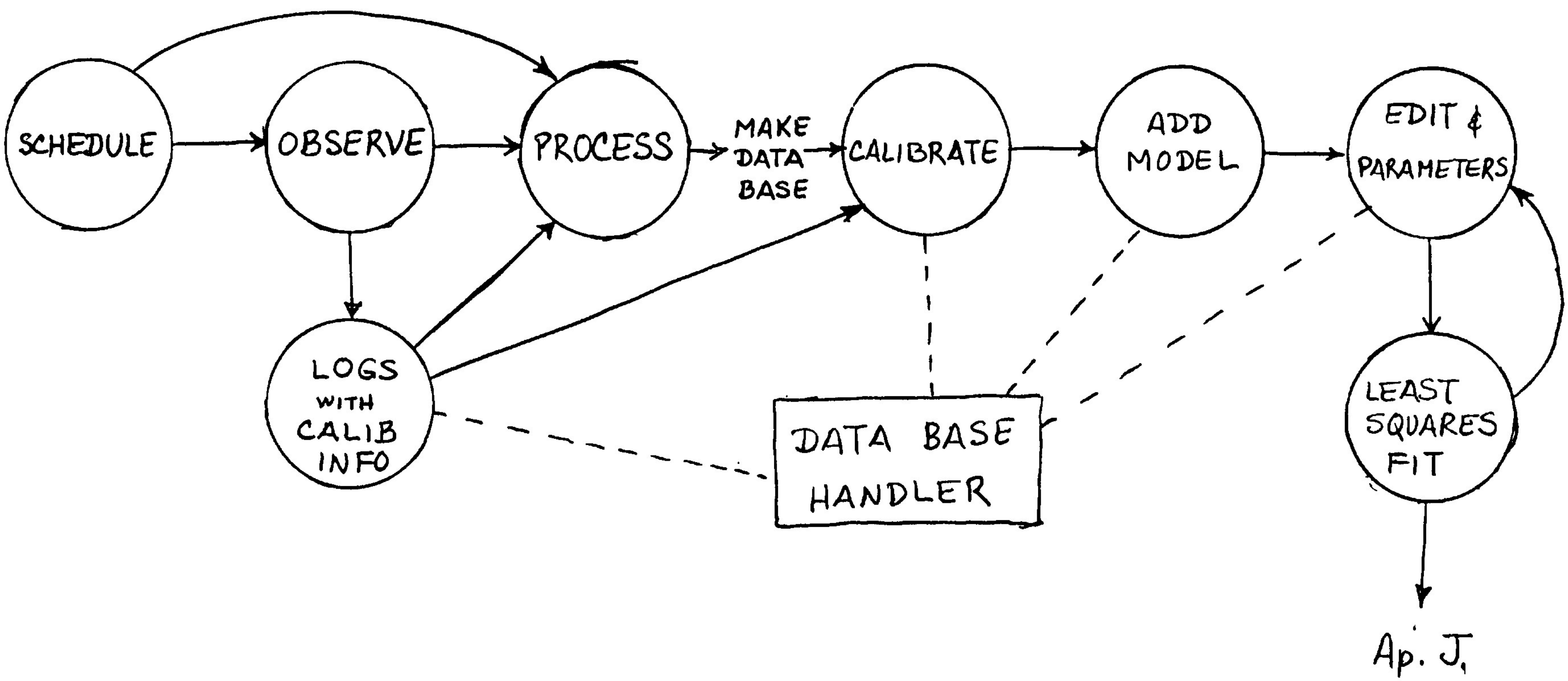
At the outset, I note that astrometric data analysis can not be separated from geodetic analysis, since precise knowledge of antenna locations and earth orientation are required for astrometry and vice-versa.

A block diagram which shows the flow of data and important analytical steps for an astrometric observation is shown on the next page. I will first discuss some of the overall design philosophy of such experiments, and then describe in some detail what goes on at each step. I will often slip into describing things as they are done at NASA/GSFC on an HP 1000 F-series mini-computer. The NASA observations are now routinely determining baselines to a few centimeters for trans- and intercontinental distances and source positions to a few milliarcseconds. A brief appendix describes some of the features of the NASA VLBI Data Base Handler, which is intimately entwined with many of the analysis steps to be described.

At the risk of boring the reader, I point out that the fundamental observables in a VLBI experiment are the amplitude and phase of the fringes, the propagation delay, and the delay rate (more often thought of as the fringe rate). Both group and phase delay can be measured. (The fringe phase is the phase delay, usually with a large integral number of turns ignored!) For a given baseline, the delay and rate values are sinusoids of 24 hour periodicity whose amplitudes and phases depend on the exact position of the source and antennas. A single astrometric observation takes a sample of these delay and rate sinusoids. A single observation may also be thought of as measuring the value and derivative of the dot product of the source and baseline vectors. Enough measurements enable the source and/or baseline vectors to be measured. A host of complications arise since the Earth is not a good inertial reference platform: it nutates and precesses and does not rotate at a uniform rate; it has tides; it has a (wet) troposphere and ionosphere; and moves in a complicated fashion with respect to the Solar System barycenter. The electronics and cabling of receivers, behavior of maser clocks, and the effects of relativity must also be considered.

VLBI ASTROMETRY DATA FLOW

(NASA MK III STYLE)



In general, these problems cause the *a priori* geometry of a VLBI system to be so imprecise that use of phase delay for measuring source positions and baselines is hopeless because of lobe ambiguity problems at wavelengths of a few centimeters. Conversely, the delay precision of a single channel is far less accurate than desired (from a few to a few tens of nanoseconds), even for bandwidths that I understand are currently being considered for the VLBA. The technique now in use is bandwidth synthesis (BWS). For BWS, several frequency channels are arrayed to synthesize a very broad bandwidth. Current techniques use 4 to 10 frequencies in zero redundancy arrays that span 100 to 350 MHz, with unit spacings of 5 or 10 MHz, and hence delay grating lobes separated by 200 or 100 nanoseconds respectively. (The missing spacings cause sidelobe levels of up to 50%, but identifying the principle lobes is no problem, and that is all that counts.) For strong sources, the BWS group delay error is around 10 picoseconds (an equivalent light travel time of only a few millimeters). Thus, potential error sources should be controlled or measured to comparable accuracy. The post-correlation fringe analysis must coherently integrate across the synthesized band for maximum sensitivity.

The effects of the Earth's troposphere currently limit the accuracy of of VLBI astrometry and geodesy. The excess zenith path delay due to the troposphere is about 7 nsec. The dry component of the troposphere can be fairly well accounted for by temperature and pressure measurements made during each observation. The wet component delay is more intractable. Current efforts to take care of the water depend on two channel radiometry done in the wings of the H O line at 1.3 cm, and relating the brightness temperatures measured at the two frequencies to path length.

The propagation delays caused by the ionosphere are removed by exploiting its 1/f dispersive effect on group delay. BWS observations at 2.3 GHz (S-Band) and 8.4 GHz (X-Band) are differenced with the appropriate weighting factors to eliminate the ionosphere delay component. This means that simultaneous dual-frequency BWS observations are a must if the VLBA is to fulfill its astrometric potential. A possible alternative to dual-frequency observations is BWS over a very broad band (>1 GHz at 8-10 GHz), and measuring phase curvature in the BWS coherent integration process. Source structure phase effects complicate this scheme, however.

Let's now look at the progress of a typical astrometric experiment (at least as now processed at Goddard).

SCHEDULING: A source whose position is to be measured must be observed several times in order to sample a range of source-baseline projections. An ideal schedule includes many interspersed observations of candidate and reference sources. In particular, observations at low elevations help to define the atmosphere, and observations covering a wide range of declinations and hour angles are needed for precise baseline determinations. Rapid slewing among sources allows monitoring of the maser frequency standard and helps identify systematic errors. Fast-moving antennas, with generous cable wrap overlap and 180 elevation coverage (i.e. they can tip over backwards) are highly desirable so that sources both to the north and south of the zenith can be observed in rapid succession.

The VLBA should have a master scheduling program that generates telescope schedules and initial file entries for all subsequent stage of analysis. We have found such a program, run interactively, to be very helpful for creating efficient schedules. The program has a catalogue of source positions and it handles all the bookkeeping necessary to control the telescopes and the Mark III recording system.

OBSERVING: As noted earlier, BWS observations at two frequencies are used. For a reasonable synthesis, a minimum of 8-10 frequency channels, split between the two frequencies is necessary. Hence, using just a few broad-band cassette recorders would seriously compromise the astrometric performance of the Array.

While the telescope (or Array) computer directs the telescope during a schedule, it must also be collecting auxiliary calibration data such as weather (temperature, pressure, humidity) for troposphere correction, water vapor radiometry, and cable length calibration. The cable calibration system continuously measures the cable length from the control room to a front-end frequency comb generator. The comb generator rail is inserted into the receiver front end and detected in the VLBI video band, either at observe time or later in the correlator. The phases of the calibrator rail are used to correct for all receiver, IF, and recorder dispersive effects. The rail is in fact the VLBI time reference. Its behavior must therefore be known to the sub-centimeter level.

The antenna itself should not contaminate the data. Deliberate deformations such as would occur in a homology design and focus motion should be known and repeatable to the sub-centimeter level. The entire antenna should be rigidly anchored to stable ground. It would be nice to know the location and expected motion of any nearby seismic areas.

PROCESSING: In processing, the several BWS channels must be coherently integrated to find the amplitude, phase, group delay, and delay rate of the fringes. The phase calibration corrections are applied to the data from each channel first. Ideally, a phase correction should exist for each frequency in the frequency domain at which fringes are treated (i.e. every 0.5 MHz if a 2 MHz channel is processed into four frequencies for fractional bit correction). This coherent search is currently done off-line at Haystack, but could be done on-line as is now being planned at JPL/Caltech. Unless the "location" of the fringes is well known before fringing, the on-line technique is not so useful because cross-correlation coefficients still need to be saved so that the data can be re-fringed as necessary. Experience will quickly lead to a good set of station and source coordinates so that most experiments will indeed know where the fringes are. New sources will be a (temporary) problem if they don't have positions better than about 1". Perhaps it should be a VLBA rule that only sources with VLA-quality positions will be processed. The processing center should keep an up to date list of positions. This same list could be used by the scheduling program.

There is some debate as to the reference point for observables that come from the correlator. For mapping, geocentric phases are convenient, so that closure phases may be formed easily. However, the "cleanest" measurements are for station referenced values. A station-referenced delay, for instance, is the difference between the clock readings noted as a given wave front passes by the stations. The time tag associated with the delay is the time at which the wavefront passed by the station chosen as the reference. The current NRAO and Caltech correlators produce (at least by the fringe stage) geocentric referenced phases. (I'm not sure about delays.) The Haystack Mark III correlator produces station oriented results. A model of the station positions and the rotation of the Earth is required to convert between these kinds of quantities. The use of such a model to change observables runs counter to the cleanest concept for astrometric and geodetic reduction: preserve the original observations and compare predictions to observations at analysis time.

CALIBRATION and MODELS: Calibration data taken during the observations, such as weather, water vapor radiometry, and cable calibration, must be applied to the data. (In fact, as alluded to just above, the observations themselves are never changed, but rather the calibrations are carried along in the analysis and applied to the predicted quantities just before comparison to the measured values.) The S- and X-Band observations are used to derive the ionosphere correction. The phase calibration was applied earlier, at the fringing step.

While the correlator model for the relative orientation of the baseline and source has to be good enough so that the fringes don't disappear, a much better model must be used for analysis. The model should predict the expected values for delay and rate. The model should follow "the rules", such as the IAU recommended treatment of precession and nutation. (To facilitate comparison of results with other observers or techniques, everybody should use the same rules.) The model must be well documented, again to facilitate comparison, or so that it can be "backed-out" later, if necessary. The likelihood that new and better models will come along, as certain processes are better understood, emphasizes the desirability of preserving raw observables and comparing them to calculated _a priori_ values, which can be changed at will and as needed.

BIH values for polar motion (X and Y values) and Earth rotation (UT1-UTC) are included in the model, although more precise values of these quantities will probably be found from the data themselves.

The derivatives of all parameters of interest are also calculated at this stage, in preparation for least-squares analysis.

Source structure must be accounted for, if necessary, by applying corrections based on source maps. A (bright?) feature in the map can be chosen for the reference point to which a position measurement applies. It is especially important to superpose the S- and X-Band maps correctly, or the ionosphere correction will cause the apparent position of time variable sources to move around.

Programs of these kinds can be run in batch mode.

DATA ANALYSIS: The first step in analysis is to remove the delay lobe ambiguities from the BWS process. With small unit frequency spacings and reasonably good knowledge of the station and source positions, this will not be much of a problem. If only a few frequency channels are available, the lobe spacings will be smaller and ambiguity resolution will be more difficult.

Some means of data editing must be available, although most bad data should have been automatically detected earlier, such as by lack of good fringes. We have found that a graphics terminal with a movable cursor is very convenient for this process. (Similar editing of amplitude data prior to mapping is extraordinarily easy compared to the shenanigans that go on in the VAX packages that I have seen!)

The data are now ready for solving for something. Again, we have found that an interactive program that gives rapid, easy access to all parameters that might be of interest is very useful. Solving for station and/or source positions or polar motion or atmosphere or clock behavior is as easy as toggling on or off the appropriate flags. Solutions may be done with either delays or rates or both. Solving for 25 to 50 parameters with 1000 observations takes a few minutes on the HP 1000. Much of that time is spent on disk access of scratch files.

A graphics terminal display of delay and rate residuals on a baseline by baseline basis has been found to be a convenient method for checking the goodness of a solution, and identifying parameters that need additional attention.

The editing, parameter setting, and least-squares fitting process go back and forth until the observer is happy with his results.

It should be possible to analyze more than one experiment at a time. Certain "global" parameters, such as station and source positions (except for radio stars!), should be determined by the whole span of data since they shouldn't change with time, whereas other parameters such as polar motion, clocks, and atmosphere will be different every day, or even within a day. Mixed-mode solutions where some parameters are fixed for the whole set of data and others are fixed only for subsets should be possible.

Appendix - The NASA VLBI Data Base Handler

The data base handler is described more fully by Ryan et al. (1980, "Radio Interferometry Techniques for Geodesy", NASA Conference Publication 2115, 337). Basically, the data base handler consists of two parts: a catalogue system, and the data bases themselves.

The catalogue system creates and tracks the location, status, and history of the data bases. A data base is created for each experiment day. (There could be a single data base for each experiment, but they would get rather large and unwieldy for multi-day experiments.) In fact, two data bases per day result because the S-Band and X-Band data streams are handled separately until the ionosphere correction is done. Data bases that are actively being analyzed are kept on disk for rapid access. Inactive data bases are archived on magnetic tape. Every time information is added, changed, or deleted, a new version of that data base is created, along with a history record that tells who did what.

The data bases themselves are named by a year, month, day, band scheme. They are not random access data bases, but the data from each sequential record may be randomly accessed after the record has been read. The data base system accesses a table of contents to know where all the variables are located within a record. A data base user need only concern himself with the variables he is interested in, and can ignore all other data from each record, and need not know anything about the format of the data records.

A typical data record contains the observables, calibrations, partial derivatives, a priori delay and rate values, and bookkeeping information for a single scan. The analysis programs access whatever of this information they need, often storing it in a scratch file for rapid access within that program. A set of header records contains information that pertains to an experiment as a whole, as well as variables such as pi or the speed of light that should be used consistently throughout a whole experiment.