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 18

Geodetic Measurements with VLBI

DAVID B. SHAFFER

RADIOMETRICS INC.

Abstract The intent of this chapter is to familiarize VLBI observers with some of the capabilities of geodetic VLBI as well as provide a brief introduction to planning and analyzing a geodetic experiment. I have assumed that in many cases the observer will be more interested in getting results than in acquiring a deep understanding of the software used and the science behind it. Somebody who is interested in becoming a geodesist, with all the understanding that implies, should attach himself/herself to one of the extant geodetic VLBI groups.

Contents

18.1 A Bit of History

It is now well established that the surface of the earth is comprised of some dozen tectonic plates which move slowly over the surface of the globe. Plate material is formed at spreading centers where two plates move away from each other, such as the mid-Atlantic ridge where magma wells up from the lower mantle. In other places, plates collide, usually forming either mountains such as the Himalayas where the Australian plate runs into the Eurasian plate or volcano chains such as the Aleutian Islands, where the Pacific plate is subducted under the North American plate. In still other places, plates slide along each other, such as the San Andreas fault in California, which marks the boundaries of the North American and Pacific plates.

The original clues to plate motion came from the geological record, such as frozen-in magnetic fields which matched across the Atlantic spreading center, as well as the long-noted "fit" (physical, as well as geological and fossil features) of South America and Africa. The inferred velocities of the plates were a few centimeters per year. Prior to the 1970's, it was impossible to measure such velocities directly by any known technique. However, with the advent of high accuracy geodetic techniques such as those of VLBI and Satellite Laser Ranging, it was quickly realized that it should be possible to measure directly the movements of the tectonic plates. Thus began NASA's efforts to develop systems capable of the centimeter accuracy required to measure plate motion in a reasonable amount of time. NASA's VLBI effort was centered at the Goddard Space Flight Center, in cooperation with the MIT/Haystack Observatory, as well as at JPL. I participated in the earliest VLBI baseline measurements of meter accuracy—which we thought was great stuff in 1971 (Cohen and Shaffer 1971)—and have been closely associated with the NASA Goddard effort since 1979. Thus, the flavor of this presentation is strongly influenced by the techniques developed by that group.

Although the geological record turned out to be a pretty good predictor of plate motion, it was not known whether the motions were episodic or smooth or how localized they were. Clearly, at some plate boundaries, the motion is highly episodic: witness the earthquakes in California associated with the San Andreas fault. Hence, the NASA program had several basic goals which evolved into the Crystal Dynamics Project (CDP) which coincided roughly with the 1980s. Measurements of the large-scale plate motions with accuracies of one centimeter per year were a prime target, along with studies of small-scale deformations that occur in seismically active areas which might precede and follow earthquakes and ultimately lead to earthquake predictions.

18.2 Size of Geodetic and Astrometric Effects

The size of tectonic motions and regional deformations is up to a few centimeters per year. There are other effects which must also be considered, in order to perform centimeter-level geodesy. To put these into perspective, it is useful to compare physical and angular displacements and the speed of light. As viewed from the center of the earth, three centimeters on the surface of the earth subtends an angle of about $0\rlap{.}''001$ (1 mas). Note also that light takes about 100 picoseconds to travel three centimeters. Thus, to measure displacements on the surface of the earth to centimeter accuracy, we must consider and measure angular effects to sub-milliarcsecond levels and worry about timing accuracies at the 10s of picoseconds level. Geodesy and astrometry are inseparable at these levels.

At these scales, the motion of the earth is far from simple. The largest effect is the precession of the earth's pole with a period of about 26,000 years, in response mostly to luni-solar gravity. Precession has the effect of sliding the right ascension/declination coordinate frame along the ecliptic at a velocity of about 50 arcsec yr^{-1}, or many milliarcseconds per day. The next largest effect is nutation, which has an amplitude of up to $9''$ and a principal period of 19 years. Polar motion, which results from the beat of several effects, one with a period of a year and one somewhat longer (400 to 440 days—the *Chandler wobble*), is another significant feature. The net effect is the displacement of the instantaneous rotational pole of the earth with respect to the earth's crust, with an amplitude of several meters over the course of a year.

The earth does not rotate terribly smoothly either. Rotational variations ("UT1") show up as fluctuations in the length of the day and irregularities in the smooth advance of sidereal time. Day-to-day effects are at the level of a few milliseconds of time, on top of a long-term slow down. At the moment, the rate difference between the earth's rotation and atomic time (UT1-UTC) accumulates at about a second of time per year, resulting in "leap seconds" which are inserted into stepped atomic time as necessary (on June 30 or December 31) to keep the two time systems within ± 0.7 second.

There are other physical displacements, too. There are tides of the solid earth. Just as the ocean responds to the differential gravitational effects of the sun and moon, so does the "solid" earth. The ground under your feet moves up and down some 30 centimeters twice per day in response to tidal forces. The displacement depends on the elasticity of rocks and the rest of the earth and is characterized by the Love numbers. As the accuracy of geodetic measurements has improved, we have been forced to consider other smaller effects. For antenna sites relatively near an ocean, there are effects of the ocean tides: the local loading of the earth varies as the amount the water changes with the tides. Even changes in atmospheric pressure have to be considered as it affects the net load experienced by the ground. Sites in northern Canada and Scandinavia are rebounding, at predicted rates of millimeters per year, since their burden from the latest ice age has been removed. Antenna structures deform at the centimeter level as they point in different directions. Although our attempts to measure the effect directly have been inconclusive, simple considerations of variations in temperature lead to predictions of changes in the effective height of these steel structures at the centimeter level. The geodetic reference point, usually taken as the intersection of the axes, is many meters above the ground. Fairbanks, Alaska undergoes temperature swings of more than 60°C from winter to summer.

All of these effects must be considered in the analysis of VLBI geodetic data.

18.3 Phase and Group Delays

In order to measure plate motions at the centimeter level with VLBI, a host of complications must be overcome. As just noted, there are many physical motions which must be considered. There are also the inevitable problems that come from making experiments in the real world!

The basic equation for VLBI geodesy (as well as astrometry) is the expression for the time delay of radio source signals received at two (or more) antennas. If $\hat{k}$ is the unit vector towards a source and $\overline{B}$ is the baseline vector pointing from the first antenna to the second antenna, the (approximate!) geometrical delay τ_g between the arrival time of signals at antenna one and antenna two is

$$\tau_g = -\frac{\hat{k} \cdot \overline{B}}{c}, \tag{18.1}$$

where c is the speed of light. By measuring τ_g for many geometries, we can solve for $\hat{k}$ and $\overline{B}$ in a least-squares sense. We can also extract useful information from the time derivative of equation 18.1, the delay rate. Besides geometry, there are other contributors to measured delays:

1. there is measurement noise that limits the precision of determining τ_g,
2. there are propagation effects through the ionosphere and atmosphere,
3. the timing standards are not perfect and also have noise,
4. a radio source may not be point-like,
5. the notation $\overline{B}$ is not quite exact—the earth moves while the signals propagate, so $\overline{B}$ is not the same when the signals arrive at the first antenna as it is when they get to the second antenna. This is not really a problem, but shows the need to be careful in how one thinks about the delay.

Let's consider exactly what kind of delay we are measuring. In a conventional interferometer, the delay is often converted to wavelengths:

$$\phi = \frac{c\tau_g}{\lambda}. \tag{18.2}$$

This is a phase delay. For baseline determination, the full count of the number of wavelengths is vital. If the a priori geometry (location of the antennas, position of the radio sources, and propagation delays) of the interferometer is not known to better than a wavelength, it is very easy to lose count of the number of wavelengths. Such "lobe ambiguities" may be difficult to detect, and they will lead to incorrect baseline and/or source position determinations. With enough phase delay measurements, one may be able to bootstrap to the correct solution by trying various additions or subtractions of ambiguities to the various measurements. As the baselines get longer and longer, this bootstrap process becomes more and more difficult, as each measurement may have a different ambiguity.

Most geodetic VLBI is carried out at GHz frequencies, or wavelengths of a few centimeters. In a VLBI system with baselines of hundreds to thousands of kilometers, the a priori geometry is generally uncertain by far more than a few centimeters. Even if the baseline length and the position of the radio sources are known this well, the orientation of the baseline due to polar motion and UT1 variations is not known well enough. Errors in estimating propagation delays through the atmosphere may also be appreciably larger than a few centimeters. Thus, most VLBI geodesy is based on group delays.

A low-precision group delay can be measured by single channel VLBI. It is the (total) delay at which the fringe amplitude is maximized. The cross-correlation amplitude as a function of relative delay between the bit streams recorded at two sites depends on the Fourier transform of the frequency bandpass. For any Fourier transform pair (such as delay and frequency considered here), the width ω of the transform in one domain is roughly $1/\omega$ in the other domain. So, a 2 MHz RF passband (the width of a Mark II or standard-speed Mark III channel) has a 0.5 μs coherence width in the time (delay) domain. For moderate SNR, the location of the delay peak can be measured to some fraction (generally about 1/SNR until systematic rather than noise errors dominate) of its width. For the single 2 MHz channel case, we might hope to achieve delay accuracies of a few to a few tens of nanoseconds. Light travels one *meter* in 3.3 nanoseconds! Hence, we can not do centimeter-level geodesy with such narrow bandwidths.

For centimeter precision on the measurement of delays, bandwidths on the order of 100 MHz are required. Conceptually, we could just sample and record a single bandwidth this wide. However, prodigious sampling rates are required: 200 megasamples per second. Although this is now technically feasible, it was not when the earliest desires to make accurate geodetic measurements arose. Just as interferometers can imitate a filled aperture antenna by sampling only part of the aperture at any one time, so it is possible to sample several narrow-band channels at widely-spaced frequencies to get the delay accuracies of a broad-bandwidth system. This combination of multiple channels to measure group delay is called *bandwidth synthesis* (BWS) (Rogers 1970), and it is the basis for all high-precision VLBI geodesy and astrometry.

There are various trade-offs to be considered when choosing a BWS scheme. They are just about the same as the trade-offs that must be considered when making an aperture synthesis map. The Fourier transform relation between antenna position (considered as a linear array) and beam pattern is exactly the same as the relation between channel frequency and the BWS delay function of fringe amplitudes. There is even diminution of the delay grating sidelobe amplitudes by the bandwidth of the individual channel, just like the primary antenna pattern reduces the sidelobes of the synthesized beam. There will be delay ambiguities at the reciprocal of the unit spacing, i.e., for a unit spacing of 10 MHz in the frequency sequence, the ambiguity spacing will be 100 ns. Since the spectra of the radio sources are featureless (except for a slope) over the BWS bandwidth, we do not have to be concerned about losing information if we don't sample all the spacings. The limit is the sidelobe level that we can tolerate, which will be set by the SNR of the individual observations, i.e., the maximum sidelobe level in the BWS function should be several sigma below the peak. If the SNR is 20, say, so that a typical fluctuation in fringe amplitude due to noise

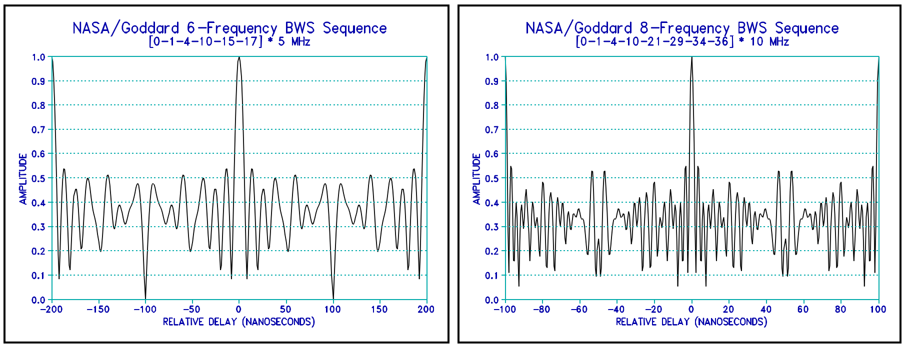

Figure 18.1: Bandwidth synthesis delay resolution functions

would be 5% of the peak, then the maximum sidelobe should probably be no more than 70% of (or 6σ below) the peak.

High sidelobe levels result when too few channels are used to span too large a frequency. The Mark III VLBI system was designed for NASA geodetic observing. It can observe 14 individual frequency channels. These channels are normally arranged as 8 channels at X-band (8–9 GHz) spanning bandwidths from 360–720 MHz, and 6 channels at S-band (about 2.3 GHz) spanning bandwidths of 85–125 MHz. The two observing bands are used to remove propagation effects caused by the ionosphere. Bandwidth synthesis delay functions are shown in figure 18.1 for these arrangements. The relative channel frequencies are shown in the figure. These are zero-redundancy sequences found by Doug Robertson (personal communication). The sidelobes are about 50%, and the ambiguities range from 50 to 200 nanoseconds depending on the frequency multiplier.

The choice of BWS sequences for the VLBA is more limited. There are a total of only eight channels available, which must be used to cover two frequency bands for ionospheric delay correction. Because of large delay variations which can be caused by the ionosphere, I do not recommend ambiguity spacings below 30 to 50 nanoseconds for S-band. For reasonable overall bandwidths (up to a few hundred MHz), this rule pretty much requires that 4 channels be used at S-band, leaving only 4 more for X-band. (At the moment, S- and X-band are the only two frequencies which can be observed simultaneously with the VLBA. You *do not* want to even think about doing precision astrometry/geodesy by switching observing bands back and forth.) The maximum length sequences for four channels with tolerable sidelobes are only 14 or 15 units long. Minimum ambiguities of 15 to 20 nanoseconds are acceptable at X-band (50 to 67 MHz unit spacing). So, the maximum frequency span at X-band can approach 1 GHz. Note, though, that 500 MHz is the maximum bandwidth which can be handled by a single downconverter. It is possible to span a wider band by using multiple downconverters, which the NASA group has done with the VLBA. It is also possible to switch the frequencies of (some of) the baseband converters to "add" additional BWS channels. However, the logistics of this are not really recommended unless you absolutely need it! The VLBA can record very broadband single channels. It may be possible to use the single channel delay, calibrated by the multiple phase-cal tones therein, to identify the correct lobe of the BWS delay, thereby allowing closer-spaced ambiguities than one would normally prefer to deal with.

The BWS channels must be properly phased for the synthesis to provide maximum sensitivity. If they are out of phase, the maximum signal will occur at the wrong delay—which would be a time bias if it stayed constant—and the sensitivity will be reduced. Generally, because of electronic variations and different paths through the system, the channels will have different and slowly varying phase offsets. Hence, a "phase calibration" signal is usually injected into the receiver, ideally at the feed but sometimes somewhere in the IF. This signal consists of 1 MHz (or sometimes 5 MHz) frequency rails, whose components are in phase when they are injected. The frequencies of the VLBA baseband converters are usually set so that these phase cal signals are recorded at 10 kHz (as well as 1.01, 2.01, etc., MHz). The correlator extracts (some of) these tones. By assuming that the tones should be in phase through all the frequency channels, phase offsets for each channel are determined and added back to the observed radio source phases. If there is no phase cal or if it fails, observations of strong sources can be used to measure directly the relative channel-to- channel phases.

The expected delay error, σ_τ, of a BWS observation is (Whitney 1974)

$$\sigma_\tau = \frac{1}{2\pi\, SNR\, \nu_{rms}}, \tag{18.3}$$

where ν_{rms} is the rms spanned bandwidth ($\nu_{rms} = \sqrt{\overline{\nu^2} - \overline{\nu}^2}$). ν_{rms} is about 0.4 times the total spanned bandwidth. For the X-band sequence shown in figure 18.1, with a total span of 360 MHz, ν_{rms} is 140 MHz. Hence, the delay error as a function of SNR is 1135/SNR picoseconds. For SNRs of 20 to 30, typical goals for a VLBI geodesy experiment, the delay errors will be 35 to 55 ps or so. This gives us the hope of doing centimeter measurements when an entire experiment is analyzed.

We discuss later how group delay ambiguities are actually resolved.

18.4 Observing Techniques and Calibrations

Geodetic observations are made in a much different style from most mapping observations. Each observation measures the baseline projection in the direction of the source (the dot product of equation 18.1). If everything else were known accurately, a few observations of a few sources would suffice to determine the baseline. However, things are not this nice. Whereas the baseline and the position of the sources can probably be considered constant for a one-day experiment (although there are sub-daily terms in polar motion and UT1), other parameters

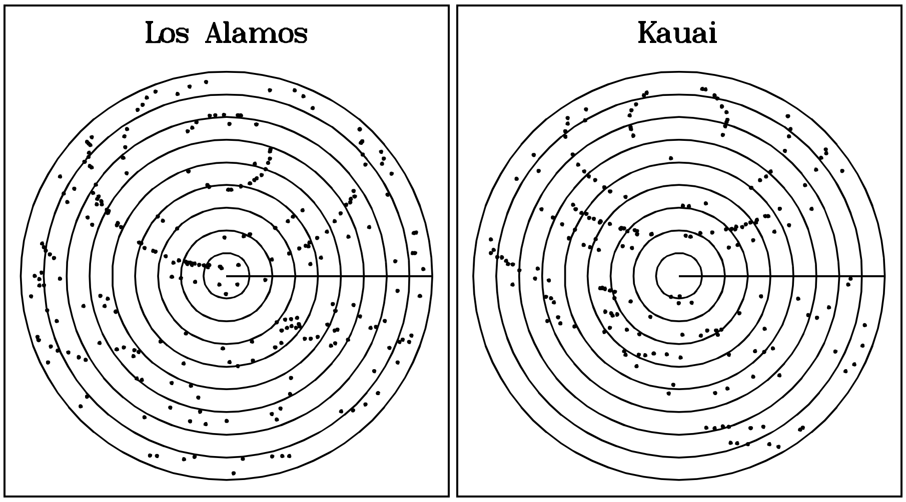

Figure 18.2: Azimuth/Elevation distributions of observations from the R&D-3 Experiment. These are "top" views of the local antenna coverage, with 90° elevation at the center.

are not. The atmosphere and clocks vary on all time scales. Our experience has shown that the best experiments are those that make as many observations as possible in as many directions at each site as possible, so we can perform a least-squares analysis.

For most arrays of antennas, it is not possible to get uniform Az/El coverage for all the antennas. The "interior" antennas can observe in all azimuths as they join antennas to the east and west, north and south. However, the antennas on the periphery of the array can not get such good coverage, e.g., if using only the VLBA, the Mauna Kea antenna can not observe far to the west because sources there will have set for the mainland antennas.

Making a schedule by hand for two or three antennas is not too difficult. One should observe sources as they rise and set, at whatever elevation limit is deemed appropriate (probably between 5° and 10°), and then make sure that the local sky at each antenna is covered fairly uniformly. We generally look at "bull's-eye plots" to visualize the az/el coverage for each site. Figure 18.2 shows the bull's-eye plots for an interior (Los Alamos) and peripheral (Kauai) antenna from a 6-station experiment.

One should choose an SNR target for the observations to ensure small-enough delay errors. For an array like the VLBA, where the antennas all have nominally the same sensitivity, you might think that the scan lengths should all be the same. However, nearly all sources will be more resolved on the longer baselines and T_{sys} can vary by a factor of two with elevation changes. Hence, you can either pick a scan length that ensures adequate SNR for the baseline where the source is most resolved or allow variable scan lengths for the various antennas. For NASA observations, the antenna sensitivities may differ by a factor of 10 within a single experiment. Hence, we use a crude model of the radio source (typically just a flux density and the size of a single gaussian component) along with the sensitivity of each antenna (including elevation effects) to calculate the scan length for each antenna for each observation. Often, the more sensitive antennas can go off and make an additional scan while the least sensitive antennas continue a long observation.

The multitude of parameters that should be considered when scheduling a number of antennas (four or more seem to be the real break point) is nearly overwhelming when it comes to making a good schedule. The best hand-schedulers spend as much or more time making such a schedule as would be required to observe it! It would take several work days to make a single 24-hour schedule. In the last few years, NASA has gone almost exclusively to the using the AUTOSKED software for making geodetic schedules. At each step of making a schedule, this program considers all possible scans, makes a covariance analysis to rank them by their contribution to the desired result, and picks the best one. A typical AUTOSKED schedule takes a few tens of minutes of CPU time to create.

Making a schedule requires selecting the sources which will be used. Ideally, there would be a large catalog of point-like sources to choose from. While sources with structure may be the bread and butter of the VLBA, they are not good geodetic reference sources. The apparent position of a source with structure will generally be a function of the baseline used to observe it. The bandwidth synthesis technique measures group delay, which is the derivative of phase with respect to frequency. A source with structure has phase variations as a function of position in (u, v) space. The BWS process can be envisioned as measurements with a congruent set of (u, v) tracks whose spacings are set by the channel frequencies. If this set of (u, v) tracks encounters the phase function of a source with structure, we can not distinguish between delay due to source position and delay due to source structure. In the worst cases—such as a nearly equal double observed at the amplitude minimum—the delay determined by BWS is just about meaningless! Delay errors of hundreds of picoseconds can occur. Although it would be possible to make corrections based on the structure, it is best to avoid this problem if you can. For NASA observing, we have chosen some 40 to 50 sources as the "best" among compact sources stronger than about 1 Jy. With the VLBA, much weaker sources could be used for geodesy. Some survey work must be done to find the best sources. The ongoing Caltech-Jodrell VLBI surveys may find these sources.

A typical experiment uses 20 to 30 sources. The sources are chosen to give good sky coverage. Each of these sources is observed on the order of ten times during a 24-hour experiment. I feel it is desirable to have multiple observations of each source. The results can then be reviewed after the experiment to check the consistency of data from each source. Theoretically, single observations of several hundred different sources would be just as good. However, it might be difficult to identify bad data in such an experiment. Most geodetic experiments run for 24 hours. This length allows separation of certain effects, such as nutation and polar motion.

The biggest error source in contemporary geodetic VLBI and the current limit on our accuracy are propagation delays. Fortunately, the effect of the ionosphere can be eliminated by observing at two widely separated frequencies. Delays due to the earth's troposphere are much harder to deal with.

The ionosphere is dispersive. Its index of refraction, and hence the propagation speed, depends very nearly on the square of the observing frequency. By taking the difference between the delays measured at two frequencies, the ionosphere's contribution to the delay can be measured very accurately. Nearly all geodetic observing uses X-band (8 to 9 GHz) and S-band (about 2.3 GHz) to make the ionosphere correction. For the ionosphere correction, the delay error at the lower frequency contributes fractionally to the overall delay error by roughly the frequency ratio squared. Thus, we use the broadest possible spanned bandwidth at the higher (X-band) frequency, where we also strive for the best SNR. At the lower frequency (S-band), more modest spanned bandwidth and SNR are adequate. If the SNRs are high enough, it is possible to use the group delay to identify the correct phase delay lobe. In this case, the group delay error at both frequencies must be small.

The atmosphere is not dispersive, so we must use other techniques to determine its contribution to the measured delays. The typical zenith path delay is on the order of 7 nanoseconds, corresponding to about 2 meters of extra path length. At low elevations, the path length, which goes as $1/\sin(\mathrm{El})$, can be much larger. The delay at a given elevation is usually modeled as the zenith delay times a mapping function which describes the path length as a function of elevation. If we are hoping for sub-centimeter results, we must characterize or measure the zenith delay to the same level or better, i.e., perhaps a part in a thousand. The mapping function for the dry component of the atmosphere is relatively well modeled, except at the lowest elevations (below 10°). The delay is just proportional to how much air there is, which is well determined by pressure and temperature measurements. We routinely measure the temperature, pressure, and humidity for each scan of a geodetic experiment.

The delay contribution due to the wet component of the atmosphere is highly variable, since it depends on the amount of water vapor in the air. The water vapor scale height is only a few kilometers, but the mapping function is not as amenable to simple models as is the dry component. Originally, it was hoped that water vapor radiometers which measure the sky brightness at 22.2 GHz (observing the same molecular transition studied in interstellar space) could determine the water vapor contribution to delay on a scan-by-scan basis. However, the care and feeding of these temperamental instruments seems to be more bother than it is worth. Absolute brightness measurements accurate to 0.5 K are required if the WVRs are to be useful. We have found that numerous low elevation scans (in the range of 5° to 20°) appear to work just as well, by allowing us to solve directly for the wet contribution to the delay. Hence, good experiment design requires adequate, rapid coverage of a wide range of elevations.

Note that by observing with good azimuth coverage, one can pretty well average out the effects of the atmosphere on the transverse components of site position, although gradients (pressure fronts) do exist. However, we can observe only those sources that are "up". Hence, vertical accuracy for site positions is worse than transverse accuracy, by about a factor of three.

The maser clocks which control VLBI data sampling are not perfect. They have both random and systematic drifts. The systematic drifts are easily modeled and solved for in our analysis. The shorter-term variations we attempt to follow by making as many scans as possible and solving explicitly for the clock behavior on time scales of half an hour to a few hours, as the density of data allow. A reference clock must be chosen at one of the sites. Typically, we deal only with clock offsets. Absolute time comparisons are much tougher. I have already mentioned the phase calibration system which allows us to phase connect the multiple channels. Another calibration system which should be in place for the highest quality geodesy is the cable length calibration system. The phase cal tones which are injected at the feed are essentially the time reference for the BWS delays. If there are changes in the length of the cable which delivers the phase cal signals to the feed, the delays will have a corresponding error. Thus, we also use a calibration system to measure the length of the cable which carries the phase cal signals. The NASA "cable-cal" system measures the phase of a standing wave on the cable. The measurements are accurate to a few mm of cable length change. (Only length changes matter. The constant length part is included in the clock offset.)

18.5 Analysis

After an experiment has been done, the tapes are correlated. This results in a collection of delays, generally both phase and group, and delay rates. There are also various ancillary calibration data (cable-cal, phase-cal, weather info, etc.). Analysis consists of calibrating the data, resolving the group delays (and perhaps the phase delays), and setting up and carrying out a least-squares solution for the parameters of interest.

I am most familiar with the flow of data for NASA experiments, and NRAO is adapting much of the NASA analysis software, so will discuss the analysis from the NASA viewpoint. The first problem to face is just keeping track of all the data. This is done in a data base system. The data base contains the rawest possible data and never modifies them. All modifications which might be applied to the data are carried along as separate values which can be added if and when desired. Also carried along for each observation are the partial derivatives which relate that scan to the parameters which we might be interested in solving for.

In particular, the NASA data bases contain the observables: total delays and delay rates, as determined by the correlator and fringe-fitting program. As long as these two tightly-coupled processes are done correctly, there is no need to remember for posterity what they did. All that matters is the result. This is unlike the mode which I think NRAO will use, which is to archive the delay and rate residuals from the correlator model. In this case, it is also necessary to remember the correlator model in order to reconstruct fully what the observables are. No matter how the observables are stored, the data base concept allows us to update our analysis models as we gain understanding of various parameters and decide perhaps to solve for additional parameters. The raw observables should always be available for the latest solution. Many of the NASA data bases have been processed using the CALC software (see below) several times. Two-decade old data are readily available for analysis with the latest models.

Equation 18.1 gives the basic VLBI delay observable. Let's look at this in more detail. We must first consider the reference frame in which we perform our analyses. We can chose a sky-based or earth-based frame. Different frames and the various effects on the observations can be related by coordinate rotations. Most of the data analysis of a geodetic experiment involves solving for the parameters of the rotations. For the moment, let us consider a sky-based frame. Radio source unit vectors are given in this frame as $(\cos\ \alpha \cos\ \delta, \sin\ \alpha \cos\ \delta, \sin\ \delta)$. There is nothing to fix the origin of right ascension except agreement on some point in the sky. The intersection of the ecliptic and the great circle of declination zero (the projection of the earth's equator) is the conventional origin of RA. However, it is very difficult to locate this to sub-milliarcsecond accuracy. Thus, conventionally a particular source or ensemble of sources are chosen to define the right ascension of the system. Declinations are absolute, at least with respect to the instantaneous pole of rotation.

On the earth, we define a right-handed coordinate frame where the principal axes are such that the x-axis and y-axis lie in the plane of the equator, with x pointing towards longitude $0°$ and y pointing towards longitude $90°$ east; and the z axis aligns with the geographic pole. (The instantaneous rotation axis of the earth is generally displaced from the pole, and small coordinate rotations, e.g., polar motion, relate the two.) Each antenna site is then defined by its (X, Y, Z) coordinates. The baseline from antenna one to antenna two is defined by the baseline vector $(B_x, B_y, B_z) = (X_2 - X_1, Y_2 - Y_1, Z_2 - Z_1)$. Longitudes on the earth are essentially arbitrary. By international agreement, the longitude of Greenwich defines the origin of the system. However, this origin has not been measured in any geodetic system. Earth latitudes are somewhat better defined, but it is really the spin radius that counts. The various practitioners of geodesy have chosen their own reference frames. The NASA system takes as a reference the position of the Haystack Observatory.

The earth rotates under the sky at the sidereal rate. We can relate earth coordinates to the sky-based frame by putting in the sidereal rotation rate ωt, where $\omega = 7.3{\cdot}10^{-}5$ rad s^{-1}. The rotating baseline has components $(B_x \cos \mathrm{GST} - B_y \sin \mathrm{GST}, B_x \sin \mathrm{GST} + B_y \cos \mathrm{GST}, B_z)$, where GST is the Greenwich Sidereal Time: $\mathrm{GST} = T_o + \omega t$ (t is time and T_o is the Greenwich sidereal time at midnight).

We can now evaluate the dot product in equation 18.1 to get

$$\begin{aligned}
-c\tau &= \hat{k} \cdot \overline{B} \\
&= B_x \cos\delta \cos(\alpha - \mathrm{GST}) + B_y \cos\delta \sin(\alpha - \mathrm{GST}) \\
&\quad + B_z \sin\delta\,.
\end{aligned} \tag{18.4}$$

along with all the nuisance delay terms which depend on atmosphere, ionosphere, clock errors, source structure, and noise. The derivative of equation 18.2 gives the delay rate (equal to the fringe rate when properly normalized). There is no B_z term in the delay rate. The behavior of the delay as a function of time shows how various parameters will affect the delay. For a particular source, the amplitude of the delay sinusoid will depend on the size of B_x and B_y and δ. The offset of the sinusoid depends on B_z and δ. The phase of the sinusoid depends on α and the relative size of B_x and B_y. By observing several sources at different declinations, it is possible to solve for the source positions and the baseline, as well as whatever other parameters put their signature in the data. A typical experiment determines a hundred or more delays, at both frequency bands, for each baseline. There may be many stations participating. The usual data processing routine is to pass the first version of the data base, which contains the correlation results, through a program which calculates the a priori expectation of what the measured delays and rates should be, based on the current best model for all parameters of interest. In the NASA version of analysis, this program is called CALC. CALC knows things like

1. the solar system ephemeris, so it can calculate the instantaneous position of the earth with respect to the solar system barycenter as well as relativistic bending due to the sun (and even Jupiter),

2. the official IAU standards for precession, nutation, earth tides, etc. in the J2000 coordinate system (Kaplan 1981),

3. models for ocean loading,

and other models or parameterization of interest. Some of the constants which CALC uses, such as π, are contained in the data base. (This promotes consistency across programs: any program which needs, say, π or the speed of light gets it from the data base, rather than the memory of whoever wrote the program!) CALC not only calculates the best guess for the a priori delay and rates, it also calculates their derivatives with respect to all parameters which might be solved for. The observables plus the derivatives are the necessary ingredients for linearized least-squares solutions for parameters of interest. All of these values are stored in a new version of the data base, which also contains all previous information.

Once CALC has been run, the geodesy experimenter is ready to run SOLVE, which is the NASA/Goddard (and adapted by NRAO) interactive least-squares solution program. SOLVE operates on the data from each baseline, but solves for station quantities such as position or clocks. The first thing that must be done with the data is to resolve the group delay BWS ambiguities. This is generally done by picking a reference clock station (they can't all be solved for, as he solution would be degenerate), parameterizing the clocks at the other sites in a simple fashion (a single offset and rate for the entire experiment), and downweighting by 10^9 the delays themselves so that the ambiguities do not confuse the solution. The delay rates are adequate to constrain the solution at this point. An examination of the delay residuals for each baseline from this first run should show strong banding about the delay ambiguities. The delays should all be moved to a single ambiguity (the one closest to zero is nice; SOLVE does this easily) for all the baselines involving the clock reference station. By doing these baselines first, and then resolving all remaining baselines to the zero delay ambiguity, you will guarantee that the clock offsets will close. If you don't have closing delays, your subsequent solutions will be garbage, as SOLVE attempts to absorb large clock offsets into all parameters!

Once the BWS ambiguities have been resolved, the ionosphere correction should be made. This requires that the ambiguities have been resolved for both the S- and X-band data. SOLVE determines the ionosphere delay correction by appropriately combining the observed delays from both bands, and puts the correction into the data base. For most experiments, this process is straightforward. Occasionally, an iteration or two may be necessary to make sure that the ambiguities have been correctly determined. If you have ignored my earlier advice to make sure that your BWS ambiguities are wide enough, especially at S-band, this is when you will discover the error of your ways. The ambiguity bands may overlap if the ionosphere has been active at one or more of your sites.

Particularly if you are observing with a new site or observing sources whose positions are not well known, you may also find that you cannot resolve the ambiguities because the geometry of the interferometer is not well enough known a priori. Remember, a delay of 50 nanoseconds could be caused by a station position error of about 15 meters or a source position error of $0\farcs5$. If this is the case, it may be necessary to make solutions using only the rates (which do not have ambiguities, but which are not sensitive to the Z-component of the baseline) or even the single-band delays (i.e., from the individual channels which comprise the BWS) which are also unambiguous but have lower precision. In this day and age of GPS receivers and source positions measured by the VLA to $0\farcs01$, ambiguity problems *should not* be a problem but they will probably turn up from time to time!

Once you are happy with the ambiguity resolution and have made the ionospheric calibration, it is time to turn the delays back on. The delays are generally far more precise than the rates, and in fact the rates are often not included in final solutions. It is then time to decide what parameters are of interest for this experiment. One commonly solves for individual station positions (but not all—there must be a reference station!), source positions, and earth orientation offsets (nutation, polar motion, and UT1). At this point, the clocks and atmospheres are generally turned on, too, using piecewise linear, continuous functions to represent clock wander and drift and atmosphere propagation variations. Typical time scales for these parameters are one to a few hours, depending on the scan density. Note that atmospheres can be solved for at all sites that are sufficiently separated.

The solving process is usually iterative. The delay (and rate) residuals should be inspected after every solution, and errant data removed from the solution. Various parameters and calibration can be turned off and on at will in SOLVE. A solution iteration takes only a minute or two, even for large experiments. Once you are satisfied with the solution, a final version of the data base should be made, which will remember all the parameterizations and editing you have made, as well as the ionosphere calibration. At this point, you can also make a “superfile”, which contains the pertinent SOLVE information for subsequent analysis. Superfiles can be retrieved later much faster than going through the entire data base file.

Superfiles are used by GLOBL, which is a superset of SOLVE for combining data from many observing sessions. The analysis of a single experiment gives you the latest position of the stations, but does not tell you how they have moved. In the good old days, site motion was studied by taking the SOLVE results and plotting them up with the results from other experiments. This gets old very quickly, and at NASA/Goddard, we can now analyze all of the experiments we have ever done (or any desired subset) for parameters which might change with time.

The parameters which affect a series of experiments can be put in two categories: “global” parameters which are applicable across all the experiments and “arc” parameters which are relevant to a single experiment. Parameters in the global category include station and site positions (perhaps with the inclusion of slow variations, i.e., the plate motion we are interested in), source positions, and the parameters of nutation, precession. Arc parameters are typified by earth-orientation parameters (polar motion and UT1) as well as the atmosphere and clocks. In some sense, these latter might also be considered nuisance parameters. In general, we don’t care what the clocks or atmosphere did. We just have to deal with them to get the information we really want.

At this time, the Goddard Mark III data bases go back to the late 1970’s. There are almost 2000 data bases and one million observations which can be analyzed together. If these were analyzed naively, the matrix inversion task would be staggering: a GLOBL run might solve for several hundred thousand parameters, most of them the nuisance arc parameters. We have developed sparse matrix techniques for GLOBL. The matrix to be inverted is fully populated only in the upper corner where the global parameter information is accumulated. The rest of the matrix consists of little islands of values along the diagonal representing the arc parameters for each experiment. By dealing with these data subsets one at a time, the matrix can be inverted in a reasonable amount of time. A typical

GLOBL run now requires several hours on our fastest computer. We now solve for tectonic rates on a global basis, and can investigate the long-term behavior of many other parameters this way.

18.6 Results

The latest NASA Goddard VLBI report presents the accumulated geodetic results of twelve years of Mark III observing (Ryan, Ma and Caprette 1993). More than 864,000 observations have been analyzed for this report, for the position of some 100 sites and 318 radio sources. Velocities are reported for many of the sites, accurate to about a millimeter per year in many cases. Most of the radio source positions are good to one milliarcsecond or better. Earth-orientation parameters (polar motion and UT1) have formal errors less than 0.2 mas and 10μs of time.

18.7 Literature

The theses by Hemenway (1974), Whitney (1974), and Ma (1978), although somewhat outdated, greatly expand upon the material presented here. In particular, Ma discusses most of the models used in CALC. The books by Meeks (1976) and Thompson, Moran and Swenson (1986) have sections devoted to VLBI geodetic/astrometric data analysis. The proceedings from IAU and other meetings present the latest details of geodetic techniques and accomplishments. See, e.g., Babcock and Wilkins (1988a); Reid and Moran (1988); Hughes, Smith and Kaplan (1991); NASA (1980); and Carter (1991).

References

Babcock, A.K., & Wilkins, G.A. (Eds.) 1988. *IAU Symp. 128, The Earth's Rotation and Reference Frames for Geodesy and Geodynamics.* Dordrecht: Kluwer Academic Publishers.

Carter, W. E. (Ed.) 1991. *Proc. AGU Chapman Conference on Geodetic VLBI, Monitoring Global Change*, Number NOS 137 in NOAA Technical Report. U. S. Department of Commerce. Rockville, MD.

Cohen, M. H., & Shaffer, D. B. 1971. "Position of radio sources from long-baseline interferometry". *Astron. J.* **76**, 91.

Hemenway, P. D. 1974. *The Measurement of Position, Baseline, and Time Using 4-Antenna Interferometry.* Ph. D. thesis, University of Virginia.

Hughes, J.A., Smith, C.A., & Kaplan, G.H. (Eds.) 1991. *IAU Colloquium 127, Reference Systems.* United States Naval Observatory. Washington.

Kaplan, G. H. (Ed.) 1981. *The IAU Resolutions on Astronomical Constants, Time Scales, and the Fundamental Reference Frame*, Volume 163 of *USNO Circ.*

Ma, C. 1978. *Very Long Baseline Interferometry Applied to Polar Motion, Relativity, and Geodesy.* Ph. D. thesis, University of Maryland.

Meeks, M. L. (Ed.) 1976. *Astrophysics: Radio Observations*, Volume 12C of *Methods of Experimental Physics.* New York: Academic Press.

NASA 1980. *Radio Interferometry Techniques for Geodesy*, Number 2115 in NASA Conf. Publ. NASA. Washington: National Aeronautics and Space Administration.

Reid, M. J., & Moran, J. M. (Eds.) 1988. *IAU Symp. 129, The Impact of VLBI on Astrophysics and Geophysics*. Dordrecht: Kluwer.

Rogers, A. E. E. 1970. "Very long baseline interferometry with large effective bandwidth for phase-delay measurements". *Radio Sci.* **5**, 1239.

Ryan, J.W., Ma, C., & Caprette, D.S. 1993. "NASA Space Geodesy Program—GSFC Data Analysis 1992". NASA Technical Memorandum 104572, Goddard Space Flight Center.

Thompson, A. R., Moran, J. M., & Swenson, G. W., Jr. 1986. *Interferometry and Synthesis in Radio Astronomy*. New York: Wiley-Interscience. First (1991) and second (1994) reprintings by Krieger Pub. Co., Malabar (Florida).

Whitney, A.R. 1974. *Precision Geodesy and Astrometry Via Very Long Baseline Interferometry*. Ph. D. thesis, Massachusetts Institute of Technology.