Lecture 2

Fourth EVN VLBI School, 3–5 Nov 1999

So You Want to Do VLBI

Bob Campbell, JIVE

Per me si va ne la città dolente

Per me si va ne l'etterno dolore

Lasciate ogne speranza, voi ch'intrate

— *D. Alighieri*

I Purpose

This introductory talk principally aims to lay the groundwork for interferometry, specifically radio-wavelength *Very Long Baseline Interferometry* (VLBI) — concepts and terminology that will recur in the lectures to follow. These will each focus in greater detail on specific areas related to the techniques of VLBI and the scientific problems that they may aid you in attacking. The present talk will cover:

- ▷ General concepts of interferometry: resolution, fringes, delays
- ▷ Observation of a point-like quasi-monochromatic source by a single two-element interferometer; how the cross-correlation function comes about
- ▷ The relationship between fringe phase and u & v; how what our arrays "see" can be related to the source's brightness distribution via Fourier transform pairs
- ▷ Problems affecting image reconstruction
- ▷ Some words about geodetic applications (there being no separate talk allotted)

I've intended to italicize newly introduced terms when they first appear, but I very well may have missed some.

This talk by itself will probably not leave you in a position to do a great deal of practical VLBI, but rather (it is hoped) to be able to get more out of later talks. I intend to leave out large areas altogether (spectroscopy, polarization, interesting points about noise analysis, *etc.*). I've intentionally tried to make these notes somewhat different from other recent "intro" talks available in proceedings from other similar VLBI schools — primarily in the increased level of detail leading up to the visibility function — which is not necessarily meant to imply that everyone will find this the best way to think about things (but hopefully there will be fewer holes to fill in yourself if you do).

I.a Some Useful References

There are several textbooks dealing with general radio astronomy, both techniques and applications:

Burke, B.F. and Graham-Smith, F. 1997, *Introduction to Radio Astronomy* (Cambridge: Cambridge University Press).

Rohlfs, K. 1986, *Tools of Radio Astronomy* (Berlin: Springer-Verlag).

Kraus, J.D. 1986, *Radio Astronomy* (Powell, OH: Cygnus-Quasar).

Verschuur, G.L. and Kellerman, K.I. (eds) 1988, *Galactic and Extragalactic Radio Astronomy* (Berlin: Springer-Verlag).

There are also textbooks dealing specifically with radio interferometry:

Thompson, A.R., Moran, J.M., and Swenson, G.W. 1986, *Interferometry and Synthesis in Radio Astronomy* (New York: John Wiley and Sons).

Wohlleben, R., Mattes, H., and Krichbaum, T. 1991, *Interferometry in Radioastronomy and Radar Techniques* (Dordrecht: Kluwer).

Meeks, M.L. (ed) 1976, *Methods of Experimental Physics*, **12C** (New York: Academic Press). Especially chapters by:

A.E.E. Rogers: *Theory of Two-Element Interferometers*, p.134

J.M. Moran: *VLBI Observations and Data Reduction*, p.228

I.I. Shapiro: *Estimation of Astrometric and Geodetic Parameters*, p.261

$\mathcal{B}$orn & Wolf (Ch. 7,10) provide the "real" background to all things interferometric, from which everything else can be derived (left as an exercise to the reader):

Born, M. and Wolf, E. 1980, *Principles of Optics* (Oxford: Pergamon Press).

$\mathcal{F}$amiliarity with Fourier transforms is a big plus, from understanding how correlators operate to getting the most out of image reconstruction, and most steps in between. Many books treat integral transforms from various perspectives (application-based, maths-for-physicists-and-engineers, rigorously mathematical), and with varying degrees of understandability (user-friendly, downright frightening).

Discrete/numerical Fourier transforms are especially pertinent; primarily mathematical texts may short-change this aspect. A small subset includes:

Bracewell, R.N. 1986, *The Fourier Transform and Its Applications*, (New York: McGraw-Hill).

Arfken, G. 1985, *Mathematical Methods for Physicists*, (Orlando: Academic Press).

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P 1992, *Numerical Recipes* (Cambridge: Cambridge University Press), Ch. 12–13.

$\mathcal{A}$ variety of proceedings of summer-schools deal with interferometry:

Felli, M. and Spencer, R. (eds) 1989, *Very Long Baseline Interferometry. Techniques and Applications* (Dordrecht: Kluwer).

Thompson, A.R. and D'Addario, L.R. (eds) 1982, *Synthesis Mapping* (Green Bank: NRAO).

Perley, R.A., Schwab, F.A., and Bridle, A.H. (eds) 1989, *Synthesis Imaging in Radio Astronomy* (San Francisco: Astronomical Society of the Pacific).

Cornwell, T.J. and Perley, R.A. (eds) 1991, *Radio Interferometry: Theory, Techniques, and Applications* (San Francisco: Astronomical Society of the Pacific).

Zensus, J.A., Diamond, P.J., and Napier, P.A. (eds) 1995, *Very Long Baseline Inter-*

ferometry and the VLBA (San Francisco: Astronomical Society of the Pacific).

There are also many symposia proceedings highlighting VLBI-driven science, including:

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

Davis, R.J. and Booth, R.S. (eds) 1993, *Sub-arcsecond Radio Astronomy* (Cambridge: Cambridge University Press).

Preuss, E. and Campbell, J. 1992, in *Lecture Notes in Physics 410: Relativistic Gravity Research*, eds. Ehlers, J. and Schäfer, G. (Berlin: Springer-Verlag), p.100: *VLBI in Astro-, Geo-, and Gravitational Physics.*

Garrett, M.A. and Colomer, F. (eds) 1997, *Proceedings of the 3rd EVN/JIVE VLBI Symposium*, *Vistas in Astronomy*, **41, 2**.

Garrett, M.A., Campbell, R.M., and Gurvits, L.I. (eds) 1999, *Proceedings of the 4th EVN/JIVE VLBI Symposium*, *New Astronomy Reviews*, **43, 8–10**.

Finally, there are innumerable more advanced papers and monographs on all aspects of VLBI and its applications. Notable among these for understanding the fundamentals of required geometric modelling are:

Sovers, O.J., Fanselow, J.L., and Jacobs, C.S. 1998, *Reviews of Modern Physics*, **70**, 1393: *Astrometry and Geodesy with Radio Interferometry: Experiments, Models, Results.* [also via the LANL pre-print server: `astro-ph/9712238`].

McCarthy, D.D. 1996, *IERS Technical Note 21: IERS Conventions.* [also via `ftp` from the U.S. Naval Observatory `ftp://maia.usno.navy.mil/conventions`.]

Seidelmann, P.K. (ed) 1992, *Explanatory Supplement to the Astronomical Almanac*, (Sausalito: University Science Books).

II General Concepts of Interferometry

The fundamental purpose of a telescope is to allow us to see fainter, smaller objects than we would otherwise be able to with the naked eye (and, of course, for radio telescopes, to extend our ability to "see" beyond the visible wavelengths). Figure 1 shows an ideal parabolic telescope (perfectly formed, having uniform unity reflectivity, sitting in a vacuum, *etc.*). A parabolic telescope will bring all points of an incoming wavefront to a single point, the focus, given that the wavefront:

i. comes from infinity (*i.e.,* a plane wavefront; parallel rays) and

ii. propagates in the direction anti-parallel to the telescope's axis.

Phase relationships among all points on the wavefront will be preserved because all paths to the focus have the same length; it is essentially this preservation of phase relationships over the wavefront that provides us the ability to measure angles on the sky. The minimum resolvable angle will scale as D^{-1}, and the overall sensitivity of the telescope will scale as the total light-collecting area, or D^2. As an example, for a telescope to "see" at the HI spin-flip transition frequency (1420.4 MHz) with the same acuity that your eye achieves in everyday life (let alone that of an optical telescope), it would need a $D \simeq 211$ m. However, construction costs will generally increase much more quickly than D^2, so the answer to fainter sources and greater resolution is not just the construction of larger telescopes.

However, we have a way out if our more pressing concern is resolution rather than sensitivity. The key point will be maintaining the phase coherence. We can remove all of the parabolic telescope's surface except for, say, the two elements drawn with thick lines in Figure 1 (in our ideal world, mechanical-engineering problems are of no concern). The rays that still reflect off these elements remain in phase as they travel to the focus, as before. Thus, if we have a partially-filled paraboloid, we can maintain the resolution we had with the entire paraboloid. We of course lose sensitivity; the total collecting area is now only the sum of the areas of the remaining two elements and the light that would have hit the area we've removed is lost. Later talks will discuss how the sensitivity of a two-element interferometer

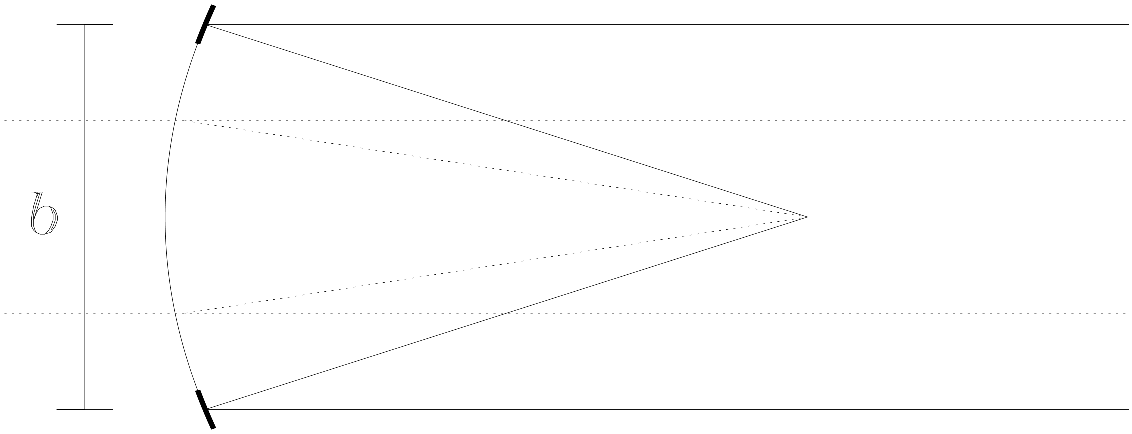


Figure 1: A parabolic telescope: transformation into an interferometer via keeping only the darkened portions of the whole parabolic surface. Solid-line rays remain in phase at the focus; dotted-line rays are lost.

scales as the geometric mean of its elements' areas ($\sqrt{A_1 A_2}$).

Again, if all we care about is preserving phase relationships among points on the wavefront, it makes no difference that our two individual elements originally came from a cut-up paraboloid or not, as long as the phase-path remains the same from each element to whatever we use as a detector. Figure 2 shows a schematic of the sort of instrument that Michelson and Pease used in the early part of this century to determine stellar diameters. A plane wavefront propagating in the direction perpendicular to the line joining the two elements (the *baseline*) will strike each element at the same time and hence the two portions will constructively interfere at the detector. If a source is observed off-axis such that the path difference from the source to each of the two elements is an integral number of wavelengths, constructive interference will still occur; conversely, if the path difference in an odd number of half-wavelengths, destructive interference will occur. Hence we have an *interferometer* — an interference-meter. Note also the parallels to a Young's two-slit experiment, where our two elements play the role of the two slits in providing two samples of a single radiation field, with a specific phase shift applied. As in the case of the two-slit experiment, we can speak of the interferometer observing a *fringe* pattern, or measuring *fringe* spacing, *etc.*

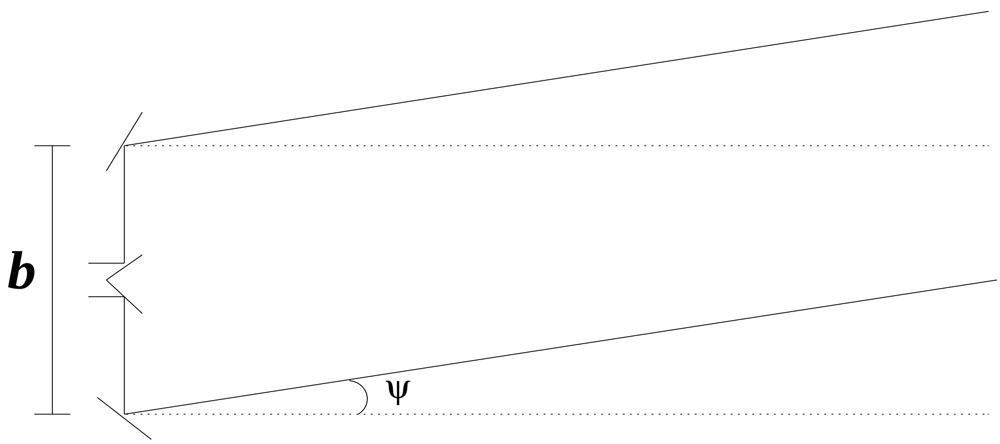


Figure 2: A Michelson-like interferometer (*i.e.,* a single baseline). Dotted lines = rays from plane wave propagating in direction perpendicular to **b**; solid lines = rays from plane wave from source moved an angle ψ off axis.

Now, if we use the interferometer (or baseline) from Figure 2 to observe some source at distance R and lying at an angle ψ from perpendicular, we can calculate the path difference via the law of cosines as:

$$\Delta R = \sqrt{R^2 + \frac{b^2}{4} + Rb\sin\psi} - \sqrt{R^2 + \frac{b^2}{4} - Rb\sin\psi}. \tag{II–1}$$

Dropping terms of order R^{-1} and below (*i.e.,* letting $R \to \infty$), we can solve for ψ:

$$\psi = \sin^{-1}\left(\frac{\Delta R}{b}\right). \tag{II–2}$$

We can plug in the criteria for maxima and minima as a function of $\Delta R/\lambda$ to compute the angular displacements of successive maxima and minima:

$$\begin{aligned} \psi_{\max} &= \sin^{-1}\left[\frac{N\lambda}{b}\right] \\ \psi_{\min} &= \sin^{-1}\left[\frac{(2N-1)\lambda}{2b}\right]. \end{aligned} \tag{II–3}$$

In real life, the tricky part is keeping the phase paths stable enough so that source-related parameters may be reliably extracted from observations of the interference fringes. "Connected-element" interferometry has to solve this problem through the use of physical phase paths (cables, waveguides, fiber optics, *etc.*) and associated engineering cleverness to calibrate instrumental effects/changes. But this poses no absolute conceptual problem. What if we use full-sized radio antennas as our elements and put them *realllllly farrrrrr* apart, say in Uzbekistan and Uruguay? Obviously, we can't physically connect two such telescopes to form the interference fringes. Rather, we need some means of recording the signal at each station for later *correlation* at a separate facility (*i.e.,* JIVE). This process will be discussed more in later talks, and the quick correlator demo will provide a feel for what happens at correlation.

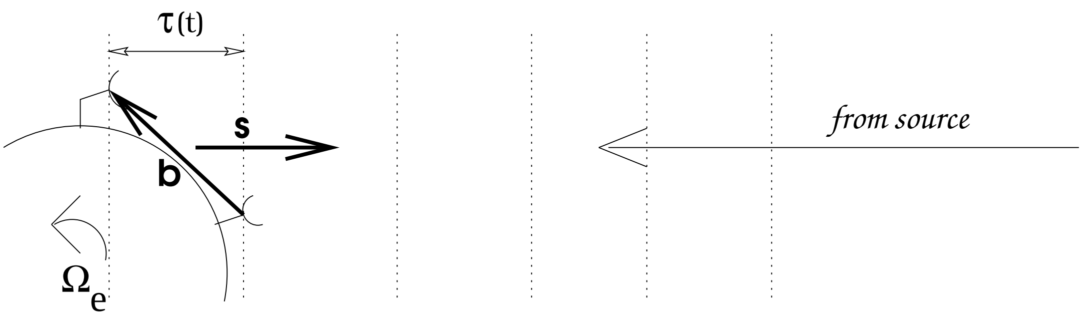


Figure 3: Radio antennas widely separated on the Earth.

Figure 3 shows two such radio antennas. As noted, there will be a *delay*, τ, in the receipt of the wavefront at one of the antennas with respect to the other. Because the Earth rotates "under" the sky, there will also be a time derivative of the delay (a *delay rate* or *fringe rate*). We need *a priori* estimates of these values prior to correlation in order to cut down the required search space and to "stop" the fringes to allow coherent averaging over a reasonable integration time, say $\sim 1\,\mathrm{s}$ (see discussion following equation III–14). The *a priori* delay and rate will depend on various geometric terms relating the orientation of the direction to the source and the baseline vector (arising from both geophysical and astronomic considerations) as well as on propagation and instrumental factors. More about this in §V.

Also note from Figure 3 that geometric component of the delay can be expressed as $c\tau = -\mathbf{b} \cdot \hat{\mathbf{s}} = -b\cos\theta$, taking θ to be the angle between the baseline vector, $\mathbf{b}$, and the direction to the source, $\hat{\mathbf{s}}$. Performing a rough variational analysis of τ and θ, we get $c\delta\tau = b\sin\theta\delta\theta$. Thus

$$\delta\theta = \frac{c\delta\tau}{b\sin\theta}, \tag{II–4}$$

and not surprisingly we obtain the best angular resolution for long baselines and when the $\sin\theta = 1$, or the source direction is perpendicular to the baseline. Note that θ as defined here and ψ as defined in Figure 2 are complementary angles.

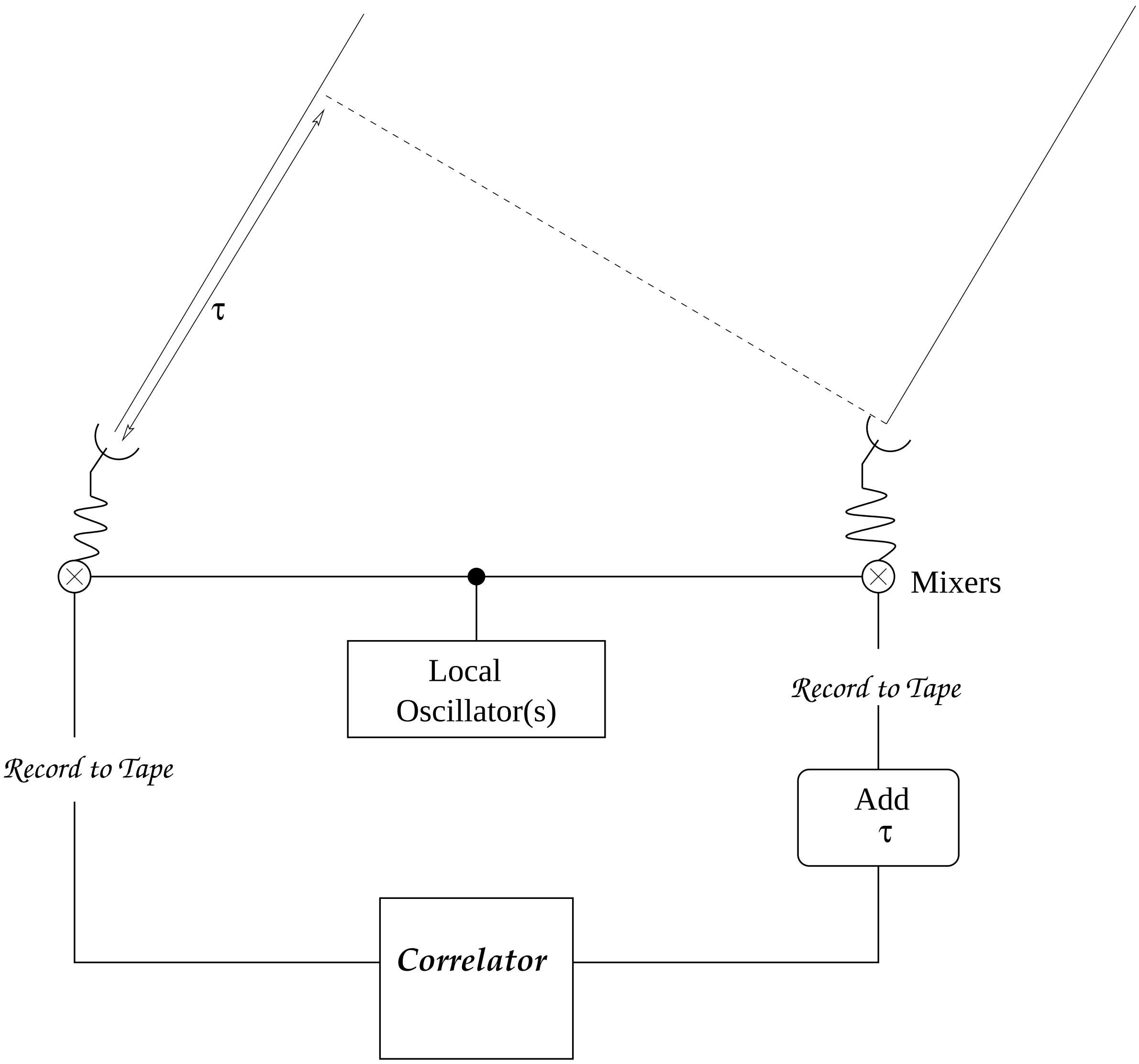


Figure 4: Simplified VLBI block diagram [level of simplification driven principally by the author's (in)ability to use `xfig`]. The addition of the delay τ to one data stream (in the rounded box) implies adding the rate $\partial\tau/\partial t$ as well.

Figure 4 shows a terribly simplified schematic of VLBI observation and correlation. The cosmic signals are received at each antenna, and are converted down to video-frequencies. This is done in stages by mixing with local oscillators (LO) that are under hydrogen-maser control.[1] These down-mixed signals are then digitized (1 or 2 bit) and sampled at a given rate and then written to magnetic tape along with accurate time marks (currently used *thin* tape is 1 inch wide and ~ 5.5 km long; *thick* tape is about half as long). The sampling rate must be at least twice the bandwidth of the channel (which typically 0.5 to 16 MHz) — this limit is called *Nyquist sampling*. The recording takes place independently at each station; the high frequency stability of the masers reduces LO noise and maintains phase coherence sufficiently to allow successful subsequent correlation (*i.e.,* reconstruction of the fringe pattern). As mentioned in the footnote, several "channels" (having different frequencies and/or different polarizations) may be recorded in parallel — more specific details about these capabilities will likely follow in subsequent lectures. At the correlator, the tapes from all the stations are played back. Values for the *a priori* delay and rate are applied to the stations' data streams (in the case of the MarkIV, to refer them to the center of the earth). The correlator then carries out the calculation of the cross-correlation function, for example as in equation (III–2).

III Coherence Function to Visibility

III.a Coherence Function to Cross-Correlation Function

Now that we've introduced the concept of a two-element interferometer, let's look at our VLBI baseline from Figure 3 in more detail. Following Born and Wolf

[1]You can see how mixing works by multiplying two sinusoids together. Let's say we want to look at a bandwidth in our cosmic signal $\omega_0 \rightarrow \omega_0 + BW$. If we multiply the cosmic signal, $\cos \omega t$, by the output of a LO, $\cos \omega_0 t$, trig identities allow us to express the product as $[\cos(\omega + \omega_0)t + \cos(\omega - \omega_0)t]$. If we low-pass filter this result to save the $\omega - \omega_0$ term, our desired bandwidth now lives in frequencies $0 \rightarrow BW$. We can have multiple LO's set to different frequencies operating in parallel. These are typically called *Video* or *Base-band Channels*.

(1980, §10.3), we can write the mutual coherence function, Γ, of the wave field from an extended source measured at points $\mathbf{r}_1$ and $\mathbf{r}_2$ (*i.e.*, the locations of the antennas forming the baseline) as:

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2; \tau) = \left\langle\, E(\mathbf{r}_1, t) E^{\dagger}(\mathbf{r}_2, t - \tau) \,\right\rangle, \tag{III–1}$$

where E is the analytic signal belonging to the (real) electric vector of light from the source at some $\mathbf{r}$, which we may characterize as a stationary, ergodic, and zero-mean Gaussian random variable; τ is the difference between the times of measurement at $\mathbf{r}_1$ and $\mathbf{r}_2$ as in Figure 3; and the brackets $\langle\ \rangle$ denote a time average. For the electric field we have described, Γ is the *cross-correlation function*, R, of $E_1(t)$ and $E_2(t)$, where the subscript denotes the point of measurement previously expressed as $\mathbf{r}_1$ and $\mathbf{r}_2$:

$$R(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} E_1(t) E_2^{\dagger}(t - \tau) dt. \tag{III–2}$$

In the simplified case of a monochromatic point-source for which we may take $E = |E| e^{i\omega t}$, we see that the cross-correlation function becomes:

$$\begin{aligned} R_{\mathrm{coh}}(\tau) &= \lim_{T\to\infty} \frac{\epsilon_0 c}{2T} \int_{-T}^{+T} |E| e^{i\omega t}\, |E| e^{-i\omega(t-\tau)}\, dt \\ &= \epsilon_0 c |E|^2 e^{i\omega\tau}, \end{aligned} \tag{III–3}$$

where we have taken the liberty of inserting the constants $\epsilon_0 c$ in order to obtain units of power per area. An actual astronomical source will have finite angular extent and will emit over a range of frequencies. Taking this to be an ensemble of independent (incoherent) emitters, we can find the total cross-correlation function of such a source observed by our two antennas by integrating the kernel of $R_{\mathrm{coh}}(\tau)$ over area on the plane of the sky and over frequency:

$$R(\tau) = \int_{\Omega} \int_{\omega} B(\alpha, \delta)\, e^{i\omega\tau(\alpha,\delta)}\, d\omega d\Omega, \tag{III–4}$$

where $B(\alpha, \delta)$ is the specific intensity (or brightness) distribution of the source ($\mathrm{W\,m^{-2}\,Hz^{-1}\,sr^{-1}}$) and $d\Omega$ is the element of spherical surface area ($\cos\delta\, d\alpha d\delta$). The integration over ω ranges over our observed bandwidth, $\Delta\omega$, centered on our reference frequency, ω_0, and taken to have a rectangular bandpass. We assume that the source's structure is independent of frequency over $\Delta\omega$, so that B can be factored from $\int_\omega d\omega$. We then have

$$R(\tau) = \int_\Omega B(\alpha, \delta) \left\{ \int_{\omega_0 - \Delta\omega/2}^{\omega_0 + \Delta\omega/2} e^{i\omega\tau(\alpha,\delta)}\, d\omega \right\} d\Omega. \qquad \text{(III–5)}$$

Evaluating the integral over ω, this reduces to

$$R(\tau) = \int_\Omega B(\alpha, \delta)\, e^{i\omega_0\tau(\alpha,\delta)} \left\{ \Delta\omega\, \mathrm{sinc}\left(\frac{\Delta\omega\tau(\alpha,\delta)}{2} \right) \right\} d\Omega. \qquad \text{(III–6)}$$

We can convert this expression to one in fringe phase, $\varphi = \omega_0\tau$:

$$R(\varphi) = \int_\Omega B(\alpha, \delta)\, e^{i\varphi(\alpha,\delta)} \left\{ \Delta\omega\, \mathrm{sinc}\left(\frac{\Delta\omega}{2\omega_0}\, \varphi(\alpha,\delta) \right) \right\} d\Omega. \qquad \text{(III–7)}$$

The *sinc* envelope in the braces decreases $R(\varphi)$ as $\varphi(\alpha, \delta)$ differs from zero. In the next section, we will introduce the concept of a phase-reference point, (α_0, δ_0), at which we define $\varphi(\alpha_0, \delta_0) \equiv 0$. As long as we stay close enough to the phase-reference point, we can replace the term in braces with an "effective" bandwidth, $\Delta\omega'$, that is essentially constant with small angular offsets from (α_0, δ_0) and can be removed from the integral:

$$R(\varphi) = \Delta\omega' \int_\Omega B(\alpha, \delta)\, e^{i\varphi(\alpha,\delta)}\, d\Omega. \qquad \text{(III–8)}$$

For examples, let's look at a longish EVN baseline (Westerbork – Noto) and an intercontinental baseline (Effelsberg – VLA, which I'll refer to as B–Y), both with a total bandwidth of 64 MHz. For the EVN baseline, the *sinc* envelope decays by only $\sim 1.6\%$ in moving 100 mas away from the phase-reference point, whereas for the B–Y baseline it decays by 5% within 42 mas. Let's say we're using a 4-MHz channel; if we look at it as a single entity, then the *sinc* envelope for the EVN baseline decays only by $\sim 0.6\%$ as you move a full $1''$ away from the phase-reference point, and for the long B–Y baseline you won't encounter the 5% decay until you are ~ 680 mas away.[2] Note that there is also a similar field-of-view limitation based on the integration time used at correlation (entering via the T in equation III–3). These are often called *smearing* effects, and they are lessened by using smaller bandwidths (*i.e.,* using many frequency channels per subband) and shorter integration times. Later talks will focus more in depth on what needs to be done if you really want to make wide-field maps.

III.b Fringe Phase in Terms of "u & v"

You hear a lot about u and v in this business — u–v plane, u–v tracks, u–v sampling (UV radiation is something completely different...). This section will try to give you three different ways to think about u and v. Let's start by obtaining an expression for the fringe phase $\varphi(\alpha, \delta)$ in terms of the baseline joining our stations. We can write the baseline vector as $\mathbf{b} = b\,\hat{\mathbf{b}}$, where the baseline unit vector is

$$\hat{\mathbf{b}}(t) \quad = \quad \cos\delta_b \cos\alpha_b(t)\,\mathbf{e}_1 \;+\; \cos\delta_b \sin\alpha_b(t)\,\mathbf{e}_2 \;+\; \sin\delta_b\,\mathbf{e}_3, \qquad \text{(III–9)}$$

and where (α_b, δ_b) is the direction in which $\mathbf{b}$ points (*i.e.,* after parallel-transporting to the center of the Earth, if convenient for visualization). The unit-vector triad $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ represents an Earth-centered, inertial right-hand coordinate system with $\mathbf{e}_3$ pointing north (*i.e,* along the Earth's rotation axis) and $\mathbf{e}_1$ pointing towards the

[2]The specific form of the *sinc* envelope resulted directly from the choice of a rectangular bandpass in equation (III–5) (the two are Fourier-transform pairs, although we did not present the development in those terms) — a different choice of filter cut-off would have resulted in some difference in the ensuing "envelope" shape.

vernal equinox (we use the basis vectors $\mathbf{e}_i$ rather than $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ since (x, y) will shortly refer to Cartesian angular displacements on the "plane of the sky"). Since the Earth rotates, the right ascension component of the baseline $\alpha_b(t)$ is a function of time:

$$\alpha_b(t) = \alpha_b(t_0) + \Omega_\oplus t. \tag{III–10}$$

(As a matter of type-setting convenience, we will generally not explicitly denote this time-dependence.) Thus, we may refer to an East-West baseline as one in which both stations are at the same latitude ($\delta_b = 0$), and to a North-South baseline as one that lies in the $\mathbf{e}_3$ direction (notice this is different from one station lying due north/south of another; the two must also be situated symmetrically about the equator — purely north-south baselines seldom have practical significance since they don't contribute much to building up u-v coverage in aperture synthesis). The unit vector in the direction of a source with coordinates (α_s, δ_s) is

$$\hat{\mathbf{s}} \quad = \quad \cos\delta_s \cos\alpha_s \, \mathbf{e}_1 \; + \; \cos\delta_s \sin\alpha_s \, \mathbf{e}_2 \; + \; \sin\delta_s \, \mathbf{e}_3, \tag{III–11}$$

where we assume the source is infinitely distant and ignore the effects of aberration.

As introduced in the discussion to Figure 3, we can express the delay, τ, in receipt of the wavefront from a distant source between the two antennas forming the baseline, and hence the fringe phase, as:

$$\begin{aligned} \tau &= -\frac{b}{c}\, \hat{\mathbf{b}} \cdot \hat{\mathbf{s}}, \\ \varphi &= -\frac{2\pi b}{\lambda}\, \hat{\mathbf{b}} \cdot \hat{\mathbf{s}}, \end{aligned} \tag{III–12}$$

using $\varphi = \omega t = \frac{2\pi c}{\lambda}\tau$ (and of course, going the other way, $\tau = \frac{\lambda}{2\pi c}\varphi$). Performing the dot product and simplifying the trig expressions yields:

$$\varphi \quad = \quad -\frac{2\pi b}{\lambda} \, \left[\cos\delta_b \cos\delta_s \cos(\alpha_b - \alpha_s) \; + \; \sin\delta_b \sin\delta_s\right]. \tag{III–13}$$

We can also get an expression for the fringe rate, keeping in mind $\alpha_b = \alpha_b(t)$:

$$\frac{\partial \varphi}{\partial t} = \frac{2\pi b \Omega_\oplus}{\lambda} \cos\delta_b \cos\delta_s \sin(\alpha_b - \alpha_s). \tag{III–14}$$

This can be up to $\sim$ 600 cycles/s for our B–Y baseline at $\lambda = 6\,\mathrm{cm}$, hence we need to apply fairly good *a priori* rates prior to correlation in order to be able to coherently integrate at all. Remember of course that φ and $\partial\varphi/\partial t$ are still functions of time via equation (III–10).

It's also sometimes more convenient to approach the situation from the viewpoint of a stationary earth, which has the sky rotating over it. In this case, the phase is:

$$\varphi = -\frac{2\pi}{\lambda} \left[(b_1 \cos H(t) - b_2 \sin H(t)) \cos\delta_s + b_3 \sin\delta_s \right], \tag{III–15}$$

and the fringe rate is:

$$\frac{\partial \varphi}{\partial t} = \frac{2\pi \Omega_\oplus}{\lambda} \left[b_1 \sin H(t) + b_2 \cos H(t) \right] \cos\delta_s. \tag{III–16}$$

In these formulae, b_1, b_2, and b_3 are the Cartesian components of the baseline in the $\mathbf{e}_i$ basis-vector frame, and $H(t) = GAST - \alpha_s$ is the Greenwich hour angle of the source, or the Greenwich apparent sidereal time minus the source's right ascension. Procedures to compute $GAST$ can be found in Sidelmann (1992) (*ch. 2*) or the *Astronomical Almanac*, among others.

Now let's expand the fringe phase about small angular displacements from an origin within the source. Letting $(\alpha, \delta) = (\alpha_0 + \Delta\alpha,\ \delta_0 + \Delta\delta)$, we can expand equation (III–13) to second order in $(\Delta\alpha, \Delta\delta)$:

$$\begin{aligned} \varphi(\alpha, \delta) &= \varphi(\alpha_0, \delta_0) + \frac{\partial \varphi}{\partial \alpha_s} \Delta\alpha + \frac{\partial \varphi}{\partial \delta_s} \Delta\delta \\ &\quad + \frac{\partial^2 \varphi}{\partial \alpha_s^2} \frac{(\Delta\alpha)^2}{2} + \frac{\partial^2 \varphi}{\partial \alpha_s \partial \delta_s} \Delta\alpha \Delta\delta + \frac{\partial^2 \varphi}{\partial \delta_s^2} \frac{(\Delta\delta)^2}{2}. \end{aligned} \tag{III–17}$$

We can define (α_0, δ_0) as the phase-reference point, and set $\varphi(\alpha_0, \delta_0) \equiv 0$.

For the remainder of the talk, we'll just worry about the first-order expansion of φ. But before we leave equation (III–17), let's just look briefly at the second-order terms to get them out of our system. These all scale as $\sim \Delta^2 b/\lambda$, where Δ stands for an angular offset in any direction. Let's look at an example: take $\Delta = 3''$ and use our B–Y baseline at $\lambda = 6\,\mathrm{cm}$ ($b/\lambda \simeq 1.33 \times 10^8$). In this case, the second-order terms contribute about $10°$ of fringe phase. Now let's say we want $30\,\mu$as angular accuracy (maybe we're trying to find any putative motion of the images in a gravitational lens system relative to each other). Our fringe spacing ($\sim \lambda/b$) is about 1.5 mas. Hence we'd like fringe phases accurate to at least $7°$. For this application, we might run into trouble without the higher-order terms.

This caveat being said, let's return to the main development. Evaluating the first-order partials of equation (III–13) yields:

$$\begin{aligned}\frac{\partial\varphi}{\partial\alpha_s} &= \frac{2\pi b}{\lambda}\cos\delta_s\cos\delta_b\sin(\alpha_b-\alpha_s)\\ \frac{\partial\varphi}{\partial\delta_s} &= \frac{2\pi b}{\lambda}\left[-\sin\delta_s\cos\delta_b\cos(\alpha_b-\alpha_s)\,+\,\sin\delta_b\cos\delta_s\right].\end{aligned} \tag{III-18}$$

We define (x, y) as Cartesian angular displacements on a plane tangent to the source and perpendicular to $\hat{\mathbf{s}}$: $x \equiv \cos\delta_s \Delta\alpha$ and $y \equiv \Delta\delta$, with (x, y) increasing in the same sense as (α, δ). We can now express the first-order expansion of the fringe phase by combining equations (III–17) and (III–18):

$$\begin{aligned}\varphi(x,y) &= \frac{2\pi b}{\lambda}\cos\delta_b\sin(\alpha_b-\alpha_s)\,x\\ &\quad+\frac{2\pi b}{\lambda}\left[-\sin\delta_s\cos\delta_b\cos(\alpha_b-\alpha_s)+\sin\delta_b\cos\delta_s\right]y.\end{aligned} \tag{III–19}$$

We can also change variables in the cross-correlation function (equation III–8) from (α, δ) to (x, y), expressing the infinitesimal surface area $d\Omega$ explicitly:

$$R(\varphi) = \Delta\omega' \iint B(x,y)\, e^{i\varphi(x,y)}\, dx dy. \tag{III–20}$$

Now let's consider the Cartesian components of the projected baseline in the directions of increasing x and y (*i.e.*, east and north). The projection of the baseline onto the tangent plane can be expressed as:

$$\mathbf{b}_p = b\,(\hat{\mathbf{s}} \times \hat{\mathbf{b}}) \times \hat{\mathbf{s}}. \tag{III–21}$$

(This is quite a handy application of the vector triple-product that you may find recurring in many situations — I can pass on from experience that you'll save time when you get to the point of remembering it rather than having to re-derive it every time it comes up). Let us define (u, v) as the easterly and northerly components of $\mathbf{b}_p$ in units of wavelengths:

$$\begin{aligned} u &\equiv \frac{\mathbf{b}_p \cdot \mathbf{e}_E}{\lambda} \\ v &\equiv \frac{\mathbf{b}_p \cdot \mathbf{e}_N}{\lambda} \\ w &\equiv \frac{\mathbf{b}_p \cdot \mathbf{e}_U}{\lambda}, \end{aligned} \tag{III-22}$$

where the unit vectors for east and north are:

$$\begin{aligned} \mathbf{e}_E &= -\sin\alpha_s\,\mathbf{e}_1 \;+\; \cos\alpha_s\,\mathbf{e}_2 \\ \mathbf{e}_N &= -\sin\delta_s\cos\alpha_s\,\mathbf{e}_1 \;-\; \sin\delta_s\sin\alpha_s\,\mathbf{e}_2 \;+\; \cos\delta_s\,\mathbf{e}_3 \\ \mathbf{e}_U &= \cos\delta_s\cos\alpha_s\,\mathbf{e}_1 \;+\; \cos\delta_s\sin\alpha_s\,\mathbf{e}_2 \;+\; \sin\delta_s\,\mathbf{e}_3. \end{aligned} \tag{III–23}$$

Combining equations (III–21) through (III-23), and after some straightforward yet toilsome evaluation of the cross-product terms, we find

$$\begin{aligned} u &= \frac{b}{\lambda}\cos\delta_b\sin(\alpha_b - \alpha_s) \\ v &= \frac{b}{\lambda}\left[-\sin\delta_s\cos\delta_b\cos(\alpha_b - \alpha_s) \;+\; \sin\delta_b\cos\delta_s\right] \\ w &= \frac{b}{\lambda}\left[\cos\delta_s\cos\delta_b\cos(\alpha_b - \alpha_s) \;+\; \sin\delta_b\sin\delta_s\right]. \end{aligned} \tag{III–24}$$

We are now in a position to interpret our sets of newly defined coordinates. As defined, (x, y) are Cartesian angular displacements on the plane of the sky tangent to the source in the directions of increasing α and δ, respectively. We can think of (u, v) in three ways.

$\hookrightarrow$ **First**, we defined them as the (x, y) components of the projected baseline (in units of λ).

$\hookrightarrow$ **Second**, we can see that u and v in equation (III–24) are, modulo a constant of 2π, just the coefficients of x and y in equation (III–19), the first-order expansion of fringe phase. Therefore, using equations (III–18) and (III–19) we may equate

$$\begin{aligned} 2\pi u &= \frac{1}{\cos\delta_s}\frac{\partial\varphi}{\partial\alpha_s} = \frac{\partial\varphi}{\partial x} \\ 2\pi v &= \frac{\partial\varphi}{\partial\delta_s} = \frac{\partial\varphi}{\partial y}. \end{aligned} \qquad \text{(III–25)}$$

The simplest solution to these simultaneous partial differential equations is

$$\varphi(x, y) = 2\pi\,(ux + vy). \qquad \text{(III–26)}$$

$\hookrightarrow$ **Third**, substituting this expression into equation (III–18), we see that the cross-correlation function becomes

$$R(u, v) = \Delta\omega' \iint B(x, y)\, e^{2\pi i(ux+vy)}\, dxdy, \qquad \text{(III–27)}$$

and thus that (u, v) is a conjugate Fourier space to (x, y). This will become the key point in the following two sections.

When multiple baselines observe a source simultaneously, they may not all lie in the same plane. We therefore need to introduce w, corresponding to $\hat{\mathbf{s}}$ in the same manner that (u, v) corresponded to $(\mathbf{e}_E, \mathbf{e}_N)$ — forming a right-handed East-

North-Up reference frame. Equation (III–25) including this "depth" term becomes:

$$R(u, v, w) = \Delta\omega' \iint \frac{B(x, y)}{\sqrt{1 - x^2 - y^2}} \, e^{2\pi i \left\{ux + vy + w\left(\sqrt{1-x^2-y^2} - 1\right)\right\}} \, dx dy. \quad \text{(III–28)}$$

One problem with w is that it destroys the simple Fourier-transform pair symmetry of $R(u, v) \rightleftharpoons B(x, y)$, as can be seen by comparing equations (III–27) and (III–28). Ignoring w will limit your field of view away from the phase-reference point in a similar manner to the bandwidth and integration-time smearing effects discussed at the end of the previous section. However, we won't discuss w further here.

Equation (III–24) provides expressions for u and v as a function of the baseline and source directions. If we re-introduce time dependence into α_b from equation (III–10), we obtain $u(t)$ and $v(t)$:

$$\begin{aligned} u(t) &= \frac{b}{\lambda} \cos\delta_b \sin(\alpha_b(t_0) - \alpha_s + \Omega_\oplus t) \\ v(t) &= \frac{b}{\lambda} \left[- \sin\delta_s \cos\delta_b \cos(\alpha_b(t_0) - \alpha_s + \Omega_\oplus t) + \sin\delta_b \cos\delta_s\right]. \end{aligned} \quad \text{(III–29)}$$

Aperture synthesis is a fancy term for of allowing the Earth to rotate under the sky, building up an elliptical track on the u–v plane for each baseline. These ellipses will be centered at

$$\begin{aligned} u &= 0 \\ v &= \frac{b}{\lambda} \sin\delta_b \cos\delta_s. \end{aligned} \quad \text{(III–30)}$$

An east-west baseline will therefore be centered at the origin. A north-south baseline will have $u(t) = 0$ and $v(t) = \cos\delta_s$, and hence doesn't contribute much to building up u–v coverage as the earth rotates under the sky. The axial ratio of the u–v track will be $\sin\delta_s$, and the semi-major axis will be $b \cos\delta_b/\lambda$. Figure 5 provides examples of full-day, uniformly sampled u–v tracks for a unit-length baseline for a number of δ_b and δ_s combinations, assuming a transparent Earth (which would indeed be a significant improvement in terms of source scheduling).

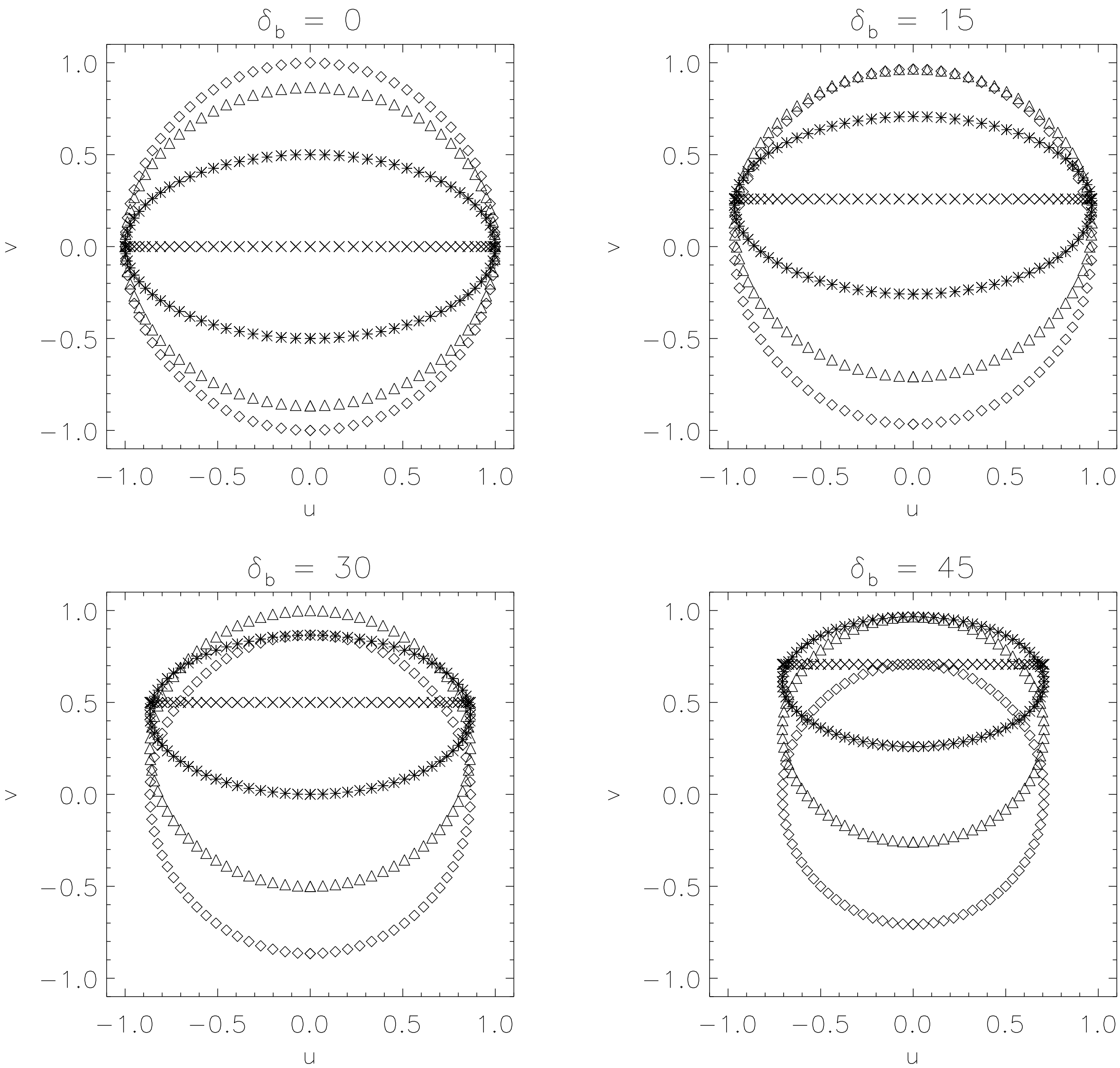


Figure 5: u–v tracks for a unit-length baseline for various δ_b. Each plot shows the results for a source at $\delta_s = 90°$ (diamond), $60°$ (triangle), $30°$ (star), and $0°$ (x). The tracks are for a full day, assuming a transparent Earth. Points are plotted every 18 min. Note the center of each elliptical track lies at $u = 0, v = \sin\delta_b \cos\delta_s$, the axial ratio is $\sin\delta_s$, and the semi-major axis is $\cos\delta_b$.

As a lead-in to mapping, we can recall that (u, v) and (x, y) form Fourier-transform pair spaces. This means that inner u–v points contain information about outer (x, y) image structure, and conversely for the outer u–v points. Gaps in the u–v coverage in specific regions will translate to more poorly constrained image information at corresponding (x, y) locations. Ideally, you'd want as full as possible coverage of the u–v plane, through some combination of many different baselines and long observing times. However, there are operational constraints on both, so you need to schedule observations to meet your scientific requirements while using the minimally necessary resources. For example, since $B(x, y)$ is a real function, its Fourier transform $R(u, v)$ will be Hermitian, which means

$$R(u, v) \quad = \quad R^{\dagger}(-u, -v), \tag{III–31}$$

where † denotes complex conjugation. Therefore, in principle you only need to sample half the u–v plane; once a u–v point is sampled, you get no new information by sampling its reflection about the origin. Consistency of u–v coverage may become an issue when multiple-epoch observations aim to investigate structural evolution within a given source.

III.c Visibility Function

Now we define the *visibility function* as the cross-correlation function of the source normalized by that of the source at zero baseline length (*i.e.*, the integral of the source's brightness distribution over solid angle, or the total flux of the source):

$$V(u, v) \; \equiv \; \frac{R(u, v)}{R(0, 0)} \; = \; \frac{\iint B(x, y)\, e^{2\pi i(ux+vy)}\, dxdy}{\iint B(x, y)\, dxdy}. \tag{III–32}$$

$V(u, v)$ is the response of the interferometer to an extended source, and may also be expressed as a phasor: $V = \mathcal{V}e^{i\phi}$, where $\mathcal{V}$ is the visibility amplitude and ϕ is the visibility phase. These expressions are also manifestations of the van Cittert – Zernike theorem, which states that the degree of coherence equals the absolute value of the normalized Fourier transform of the intensity function (*i.e.*, brightness distribution) of the source. Since $V(u, v)$ is merely $R(u, v)$ scaled by a constant, it retains Hermitian properties.

As hinted to in the previous section, aperture synthesis and image reconstruction (to be discussed presently) work by allowing the Earth to rotate beneath the source, with each baseline of fixed antennas sampling an elliptical track of (u, v) points (interrupted, of course, physically when the source descends below the horizon at either of the antennas, and also operationally by the need to change tapes, observe calibrators, allow others to observe their experiments, *etc.*).

IV Image Reconstruction

So if we stay within the assumptions that our source is incoherent (can be treated as an assembly of independent radiators), that it lies in the far-field so that we have a plane wavefront at Earth ($R \gg b^2/\lambda$, or $\gg 0.035\,\mathrm{pc}$ for our B–Y baseline at $\lambda = 6\,\mathrm{cm}$), and that the first-order expansion of fringe phase and other time- and bandwidth-smearing effects pose no problem, then all we have to do is take the inverse Fourier transform of $V(u, v)$ to come up with the source's brightness distribution $B(x, y)$.

But not so fast. We only measure the amplitude and phase of the visibility function at a finite set of (u, v) points, such that

$$V_{\mathrm{obs}}(u, v) = V(u, v)\, W_{\mathrm{obs}}(u, v). \tag{IV–1}$$

We typically take $W_{\mathrm{obs}}(u, v)$ to be unity for sampled (u, v) points and zero elsewhere (so-called *natural weighting*); one could also weight W by the anticipated baseline sensitivities. Another common tactic is *uniform weighting*, which essentially decreases the weight of each individual u–v point in proportion to the total number of u–v points in its vicinity. By the convolution theorem, which states that the Fourier transform of a product of two functions is the convolution of the Fourier transforms of each, we get a parallel expression

$$B_{\mathrm{obs}}(x, y) = B(x, y) \otimes w_{\mathrm{obs}}(x, y). \tag{IV–2}$$

Knowing only V_{obs} and W_{obs}, we want to get B. Two problems immediately stand in the way:

▷ We just can't divide through by the W's/w's because of the zeroes of these functions. This goes beyond the astronomy; we just have an under-determined inversion (sometimes referred to as a deconvolution problem).

▷ The phases of individual $V_{\rm obs}$ measurements may be corrupted by various propagation and instrumental effects (and depending on the quality of the antenna calibrations, maybe the amplitudes are too).

Image reconstruction deals with how to get around these two problems, and is the subject of a good fraction of the lectures. Here we'll provide just the quickest of overviews, somewhat artificially treating each problem separately in turn.

Let's first assume that our array measures the amplitude and phase of $V_{\rm obs}$ perfectly at the sampled u–v points. Visibilities at unsampled (u, v) points are clearly not constrained by observations, and thus there may be multiple derived brightness distributions that are consistent with the data we actually have. The problem is how to pick the best B from among this set — or perhaps at a lower level, how even to define "best". If we simply take the inverse Fourier transform of $V_{\rm obs}$ (*i.e.*, assume all unmeasured u–v points have $V = 0$), we will wind up with a brightness distribution that has a great deal of oscillation on small angular scales, and will even be less than 0 in places. This is clearly non-physical, and we need to do better. By way of terminology, however, this immediate transform is often called the *dirty map*; similarly, the transform of $W_{\rm obs}$ is called the *dirty beam*:

$$\begin{aligned} D.M. \quad &\rightleftharpoons \quad V_{\rm obs} \\ D.B. \quad &\rightleftharpoons \quad W_{\rm obs}. \end{aligned} \tag{IV–3}$$

The dirty beam is the response of our array to a point source (the point-spread function in optical-telescope parlance). If the dirty map looks just like the dirty beam, you've got yourself an unresolved source.

We can take care of some of the under-determinacy of our image reconstruction problem by making assumptions about our source and turning them into constraints on B in the image plane. If we require that B comprises a set of components, each of which has an analytic Fourier transform (*e.g.,* optically thick elliptical Gaussians, optically thin spherical shells, *etc.*), we could estimate the parameters of the components via non-linear least-squares fitting. Nothing in nature really is an elliptical Gaussian, of course, but this method does provide at least formal uncertainties in and correlations among parameters defining the orientation/shape of the components. You should always check the observed u–v visibilities against values calculated from the final model to judge whether such morphologically simple models adequately represent the underlying observables. In the other extreme, you should also perform F-tests when adding additional elliptical Gaussian components, to make sure the underlying observables can really support them. Model-fitting on the u–v plane can also incorporate additional image-plane constraints if applicable (say, a magnification matrix relating corresponding components in different images of a gravitational lens system).

The `CLEAN` algorithm assembles a brightness distribution for our source from a set of point sources placed at various locations in the x–y plane. It essentially puts a point source at (x, y) corresponding to the maximum of the dirty map (with amplitude cut down by some gain parameter), subtracts a gain-normalized copy of the dirty beam centered at this (x, y) from the u–v plane, and iterates. Hopefully, the flux of the point sources (*CLEAN components*) will converge gracefully to the noise level. (If you want, and `CLEAN` model is just an elliptical Gaussian model with *many* components, each of which is a point source — the simplest of all elliptical Gaussians.) The Maximum Entropy Method (MEM) makes a different sort of *a priori* image constraint. It tries to return the smoothest image possible (keeping $B > 0$) in the sense of having minimum excursions from the mean B everywhere over the image. In contrast to `CLEAN`, MEM can handle extended image structure better than strong point sources. This raises the possibility of combining the two methods for complicated sources. In any case, you'll still want to look at the observed–model residuals in the u–v plane as mentioned above.

But the under-determinacy of V_{obs} is not our only problem. The observed phase of each visibility point is not useful directly in constraining the image reconstruction, and depending on the quality of the station calibrations, there may also be biases in the amplitudes. A rough check on the consistency of the amplitude calibration can come from looking at "u–v crossings", individual u–v points from different baselines that happen to lie very close to each other. Under the assumption that $V(u,v)$ is not varying as a function of time over the observations, the amplitudes at these points should be about the same.

Closure phase, ϕ_c, is the sum of simultaneously observed phases of a source on three baselines forming a triangle: $\phi_c(t) = \phi_{AB}(t) + \phi_{BC}(t) + \phi_{CA}(t)$ for stations A, B, and C, for which the correlator applies the relevant delay rate terms to retard/advance the phase measurements on all three baselines to a common epoch, if necessary. The advantage of closure phase is that it is insensitive to station-specific perturbations such as propagation or instrumental effects but retains information about baseline-based effects — predominantly source structure. Closure phases can be used in place of individual phases as constraints to non-linear least-squares fitting of elliptical Gaussian models or in reconstructing a model-dependent set of phases during "hybrid mapping" iterations (see below).

If you have a reference source lying nearby on the plane of the sky, difference phase, $\phi_d(t) = \phi_1(t) - \phi_2(t)$ for the two sources 1 and 2, are also much less sensitive to propagation and instrumental effects. This concept forms the basis for relative astrometry and phase-reference mapping procedures. Here, the cycle time between the sources and their angular separation on the sky should both be small compared to temporal/spatial gradients of the propagation and instumental effects in question.

Most maps nowadays are created using a model-based, iterative scheme for determining amplitude/phase solutions consistent with the constraints derivable from the observed visibility points. For instance, even though the individual phases may be corrupt, the closure phases are still good constraints on the image structure. However, given N stations, the number of independent closure phases at any instant $((N-1)(N-2)/2)$ is less than the number of individual phases $(N(N-1)/2)$. A source model (say a point source to begin with) can be used to calculate enough "trial" phases to make up the shortfall. The source is then `CLEAN`ed (or fit with elliptical Gaussians), resulting in a new source model (set of `CLEAN` components/elliptical Gaussians). This model is then used to drive the next iteration of calculating a set of trial phases and `CLEAN`ing. Station sensitivities (scaling the amplitudes) can also be adjusted by a similar process. If all goes well, the resulting series of $B(x, y)$ from these iterations should converge (differences in structure between iterations $\rightarrow 0$, noise level in the map levels off, station-gain corrections $\rightarrow 1$). Depending on how/where one wants to draw lines between definitions, these sorts of procedures go by the names *self-calibration* or *hybrid mapping*. One way to think about things is to let $V(u, v)$ be replaced by $G_1 G_2^\dagger V(u, v)$, where the factors G_i are (time-dependent) complex station gains, which must also be estimated in a way consistent with the measured $V_{\rm obs}(u, v)$ when computing $B(x, y)$.

As with any under-determined system where *a priori* assumptions about the result (*i.e.,* map) provide the missing constraints, you need to exercise some caution to make sure you're not fooling yourself. Within the typical `CLEAN`-based self-cal/hybrid-mapping scheme, it's possible to bias the final map towards the assumed model or other undesired outcomes if you're being inattentive (or towards *a priori* desired outcomes, I suppose, if you're being devious). For instance, if you start with a very short amplitude self-cal solution interval and very small `CLEAN` boxes tightly mimicking the *a priori* model, your final maps will bear an uncanny resemblance to your *a priori* model. The basic caveat: avoid using operations meant for fine-tuning too early in the game. There are also fairly well-known intermediate stages to watch out for, such as spurious symmetrization, which will likely be discussed in subsequent lectures. Special care must also be taken in cases where the source's brightness distribution may change significantly during the course of your observation at a single epoch (*i.e.,* as you're building up u–v sampling), as could happen for Intra-Day Variable sources and/or extreme scattering events in the ISM.

V A Few Thoughts About Geodesy

We have briefly discussed the *a priori* delay and rate used in correlation. Typically, we try to get these a close as possible by using a model that describes the orientation of the baseline(s) with respect to the source(s) at the time(s) of observation, as well as other propagation and instrumental factors affecting receipt of the cosmic signals. An astronomer will in turn typically use phases "residual" to this model in subsequent mapping analysis. Now, it is possible to invert this typical situation: by observing a variety of sources and using the total delays and rates (and maybe even phases) from the correlator, we can investigate geophysical processes that enter into the model itself. Let's take a little time to look at some of these applications, without worrying about the practical differences in observation/data-reduction strategies. For those interested, Sovers, Fanselow, and Jacobs (1998) review and the IERS conventions listed in §I.b provide details of the modeling.

The total observed delay for a given baseline can be broken up conceptually into individual components:

$$\tau_{\rm obs} = \tau_{\rm geom} + \tau_{\rm str} + \tau_{\rm trop} + \tau_{\rm ion} + \tau_{\rm inst} + \tau_{\rm noise}. \qquad \text{(V–1)}$$

$\tau_{\rm geom}$ is the thing we're really interested; let's quickly first take care of the others.

The structure contribution about a phase-reference point, $\tau_{\rm str}$, follows from the discussion in §II.b. The structure of a source with respect to the phase-reference point gives rise to a phase term, the argument of the complex visibility, that in turn results in a corresponding structure delay: $\tau_{\rm str} = \frac{1}{2\pi\omega}\arg\left(V(u,v)\right)$. We can calibrate for $\tau_{\rm str}$ by mapping the source, and then computing the visibility phases that such a $B(x,y)$ would yield at our specific (u,v) coverage.

The wavefront from a source is distorted by spatial and temporal variations in the index of refraction in those atmospheric parcels through which it passes. Within the neutral atmosphere (most significantly in the troposphere), the index of refraction changes because of the induced and permanent dipole moments of the constituent molecules, including water vapor, which are functions of temperature and the partial pressures of dry air and vapor. The physics of these processes is well understood; the problem lies in the spatial/temporal distribution of individual species. We typically break the atmosphere into a *dry* and a *wet* component (*i.e.,* water vapor in the latter, and everything else in the former), and use a zenith delay plus an elevation-based mapping function for each station in order to compute dry and wet delays for an arbitrary line of sight. The fundamental problem lies in the wet component. Its distribution is highly clumpy in both time and space, and is characterized by very incomplete mixing so that surface measurements aren't really a sufficient proxy. Measurements by water-vapor radiometers or occultations of low-Earth orbiting satellites with an on-board GPS receiver may hold promise for line-of-sight wet-component delay estimation. Within the ionosphere, the index of refraction becomes frequency-dependent owing to the presence of free electrons, scaling to first order as $1/\omega^2$. Dual-frequency observations can use this dispersiveness to estimate the delay at either frequency. Single-frequency observations are more problematic, especially $\lesssim 2\,\mathrm{GHz}$. See the contribution by Campbell in Garrett *et al.* (1999) for one approach to ionospheric calibration.

Each station may also have an instrumental phase shift arising in the signal path from the antenna feed to the video-frequency signal sampling. Contributors to this term may include effects from polarization impurities in the antenna or feed or instabilities in the frequency standards in the local oscillator chains. These phase shifts would combine to create instrumental delays on the various baselines. Each measurement of the observed delay also has an associated noise term.

Finally, we can express the geometric delay in an Earth-fixed parameterization via equation (III–15):

$$\tau_{\mathrm{geom}}(t) \quad = \quad -\frac{1}{c}\ \left[(b_1 \cos H(t) - b_2 \sin H(t)) \cos\delta_s + b_3 \sin\delta_s\right], \qquad \text{(IV–2)}$$

Thus, τ_{geom} provides a handle to investigate geophysical parameters that may affect $\mathbf{b}(t)$ in the celestial reference frame. We can roughly categorize these in terms of processes that affect the Earth's rotation, orientation, or shape.

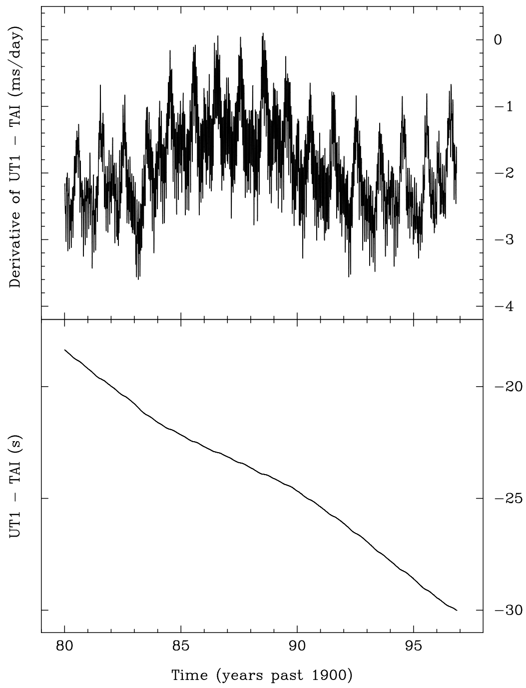


Figure 6: Observed rotation rate of the Earth from 1980–1996. The bottom plot shows the trend of time measured by the Earth's rotation (UT1) with respect to atomic time (TAI). The top plot shows the increase in the length of day, currently around 2 ms per day. Figure from Sovers, Fanselow, and Jacobs (1998).

Earth's rotation rate shows variations on virtually all time scales. Relevant physical processes include a secular increase in the Earth-moon distance, tidal dissipation of rotational energy, frictional and electro-magnetic damping at the Earth's core-mantle boundary, and even frictional forces between winds and land topography. Figure 6 shows the observed behavior of the Earth's rotation over the last 16 yr. The current trend of $\sim +2$ ms/day (UT1 slowing down $\Rightarrow$ rotation slower $\Rightarrow$ day longer) is considerably greater than the geological time-scale trend: paleological tidal sediment studies suggest that a day was 18 hr 900 Myr ago (a net $\sim +0.7\,\mu$s/day). There is evidence that the slope of the UT1–TAI curve has reversed sign in the last ~ 100 years. Effects at the core-mantle interface can provide a signal of 4–5 ms over time scales of a few decades. The annular and shorter-term behavior of the rotation rate is primarily driven by seasonal variations in global winds. Thus VLBI can have applications to medium-term weather/climatic forecasting via monitoring atmospheric angular momentum (including the ability to detect El Niño).

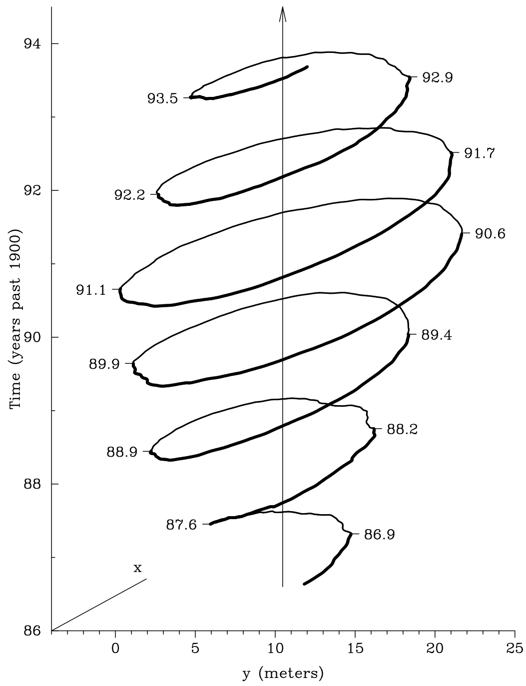


Figure 7: Observed polar motion from 1986–1993. This figure is also from Sovers *et al.* (1998).

Variations in the orientation of the Earth's rotation axis with respect to both the crust and inertial space is largely an effect of the non-spherical Earth. Change in the orientation of the rotation axis with respect to the crust is termed *polar motion*. It has a component with a period of about a year due to the same sort of atmosphere-surface interaction mentioned above and another with a period of $\simeq$ 433 days corresponding to a free-oscillation mode of the solid Earth (the Chandler wobble). Figure 7 shows polar motion in x and y from 1986–1993, with time as the third axis. The radius of the total oscillation ranges from ~5–10 m (or about 150–300 mas as seen from the center of the Earth). Changes in the orientation of the rotation axis with respect to inertial space derive from gravitional/tidal torques from the sun, moon, and other solar-system bodies on the non-spherical earth. Solar and lunar torques cause the rotation pole to precess around the ecliptic with a period of 25,800 years ($\sim 50''$ per year). The plane of the ecliptic also precesses somewhat, primarily due to the effect of the other planets: the vernal equinox moves eastward at $\simeq 12\overset{''}{.}5$ per century and the obliquity of the ecliptic (now $\simeq 23\overset{\circ}{.}5$) decreases by $\simeq 47''$ per century).

Periodic tidal effects on top of the bulk precession are collectively termed *nutation*. The principal causes are the varying locations of the sun and the moon with respect to the $\simeq 23\overset{\circ}{.}5$ tilt of Earth's equator with respect to the ecliptic and the $\sim 5^\circ$ inclination between the plane of the lunar orbit and the ecliptic. The IAU 1980 Theory of Nutation calculates instantaneous nutation from a series of harmonic terms ($N = 106$); the dominant term has a period of 18.6 yr (close to the saros cycle used in eclipse prediction) and amplitudes of $17\overset{\prime\prime}{.}2$ in longitude and $9\overset{\prime\prime}{.}2$ in obliquity. However, VLBI observations over the past two decades have clearly shown this is not good enough. Figure 8 plots the discrepancies between VLBI-derived nutation and the IAU 1980 model. Both long- and short-term deviations can be seen. Development of a newer nutation model has been an important focus recently. The IERS 1996 Theory of Precession/Nutation expands the harmonic series to $N = 263$, with periods of individual terms ranging from $\sim$5 days to $\sim$180 years, and amplitudes as low as a few μas.

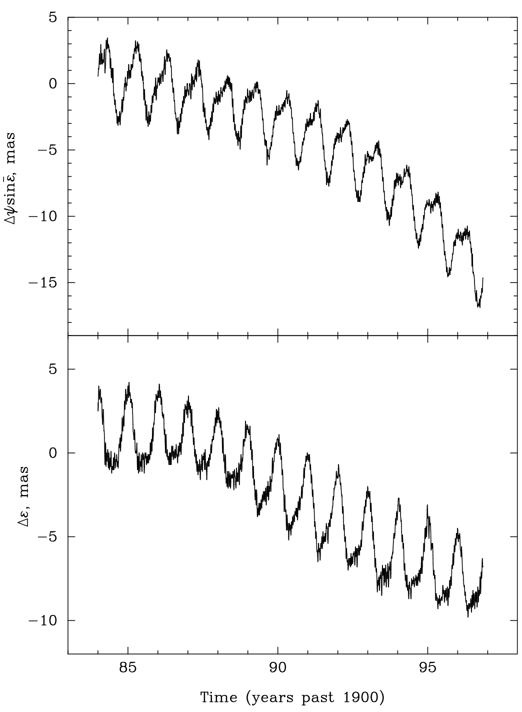


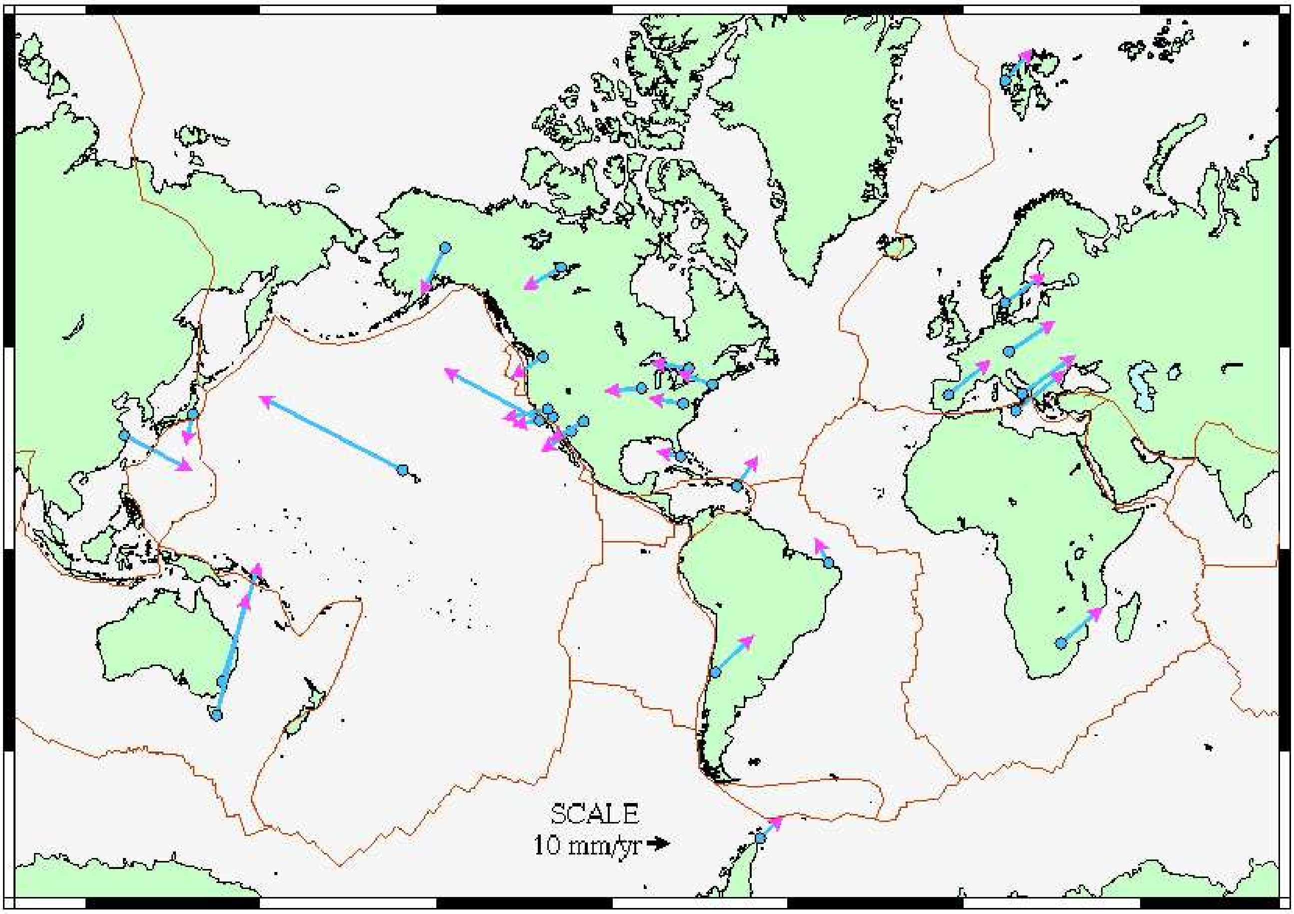
Figure 8: Discrepancies between observed nutation and the IAU 1980 model from 1984–1996 (top=longitude; bottom=obliquity). This figure is also from Sovers *et al.* (1998).

Baseline vectors will also change detectably when the figure of the Earth changes or as plate tectonics move land masses with respect to each other. The principal tidal effect is the solid-earth tide in response to the differential gravitational fields of the sun and moon. This works the same as regular ocean tides, but of course has a smaller and more complicated response because of the non-isotropic distribution of rock elasticity. Typical amplitudes are about 30 cm. There are also four other, smaller responses:

- Pole tides — the elastic response of the Earth's crust to shifts in the orientation of the rotation axis (*i.e.,* polar motion), providing station movement of ∼1–2 cm
- Ocean loading — crust deformation due to the shifting weights of ocean water from ocean tides (most important for coastal stations), providing station movement of a few cm. This shifting of ocean waters can also affect UT1 and polar motion, as previously discussed.
- Atmospheric loading — the (largely) seasonal pattern of climatic high- and low-pressure areas passing over a station, amounting to several mm.
- Post-glacial rebound — about 10 kyr ago, much of Europe was covered by an ice sheet. As the ice melted and the water ran off, there was a relatively sudden shift in crust loading. VLBI can detect the slow uplift of the crust in response to this as a still ongoing vertical relaxation of up to several mm per year.

Motions of individual stations, deduced from time-dependent baseline vectors, can be estimated with uncertainties below 1 mm/yr with the nearly 20 years of geodetic observations. Figure 9 shows selected station velocities from the Goddard Space Flight Center terrestrial reference frame solution 1102g from August 1998. (All the figures related to station motion have been downloaded from `http://lupus.gsfc.nasa.gov/vlbi.html`.)

Figures 10–12 show station velocity components in an Up–East–North reference frame for Mojave (southern California, east of the San Andreas fault), Vandenberg (southern California, west of the San Andreas fault), and Onsala, respectively. The velocity components from this GSFC solution 1102h2 are tabulated below:



NUVEL1A-NNR reference frame.

Figure 9: Selected station velocities (GSFC solution 1102g4), showing some plate boundaries.

Station	$V_{\rm east}$	$V_{\rm north}$	$V_{\rm up}$
Mojave	-17.54 ± 0.08	-3.10 ± 0.08	-1.69 ± 0.24
Vandenberg	-43.00 ± 0.23	$+23.54 \pm 0.25$	$+1.51 \pm 1.24$
Onsala_{60}	$+17.61 \pm 0.06$	$+12.98 \pm 0.06$	$+3.13 \pm 0.21$

Comparing Figures 10 and 11 sheds light on the existence of strains along the approximately NW–SE San Andreas fault (and the resulting occasional earthquake), with the Pacific plate moving northwest $\sim$40 mm/yr with respect to the North American plate. Figure 12 shows some degree of post-glacial rebound at Onsala (in the top panel of "Up" motion).

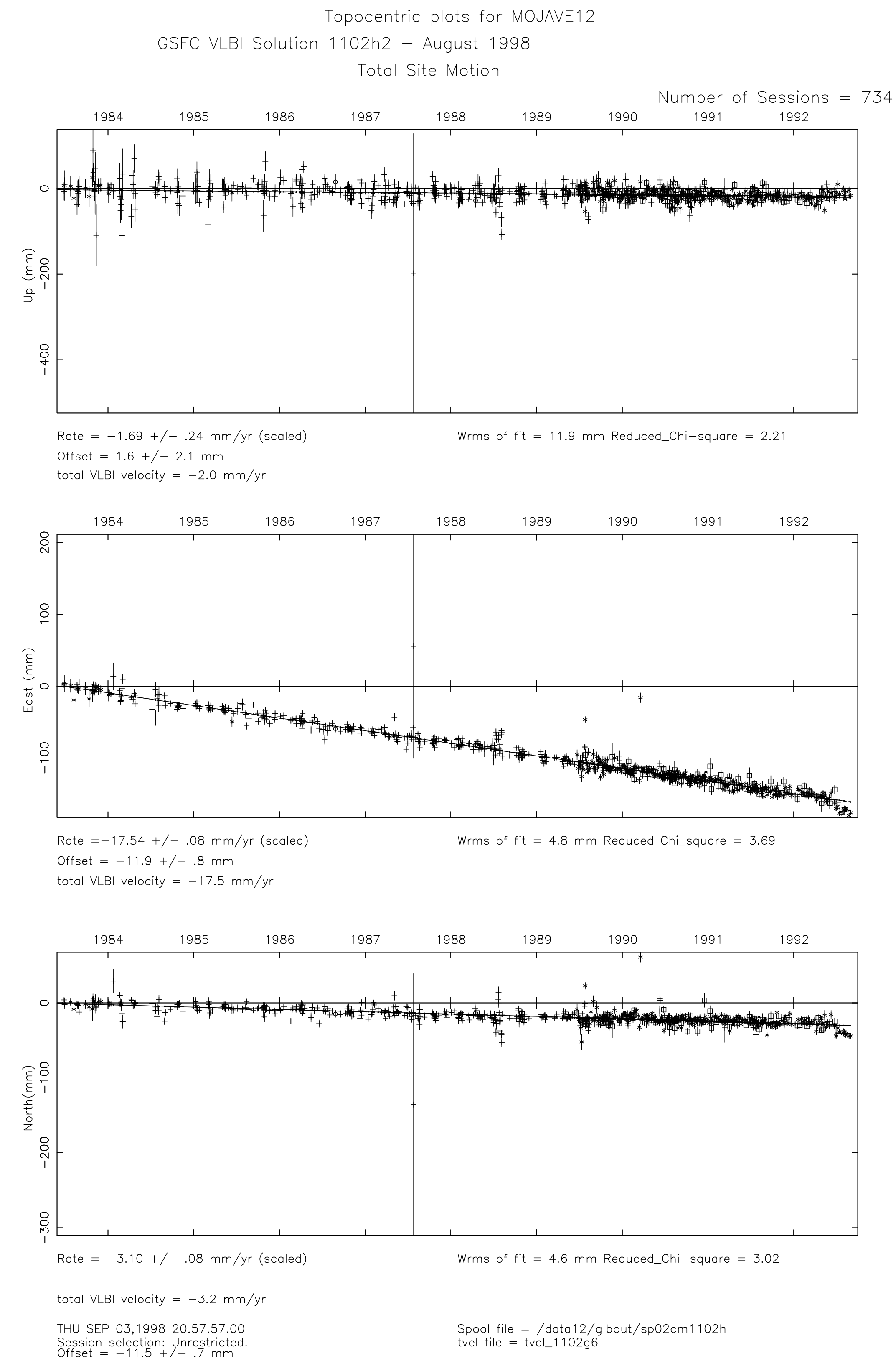


Figure 10: Station velocity components for Mojave (GSFC solution 1102h2).

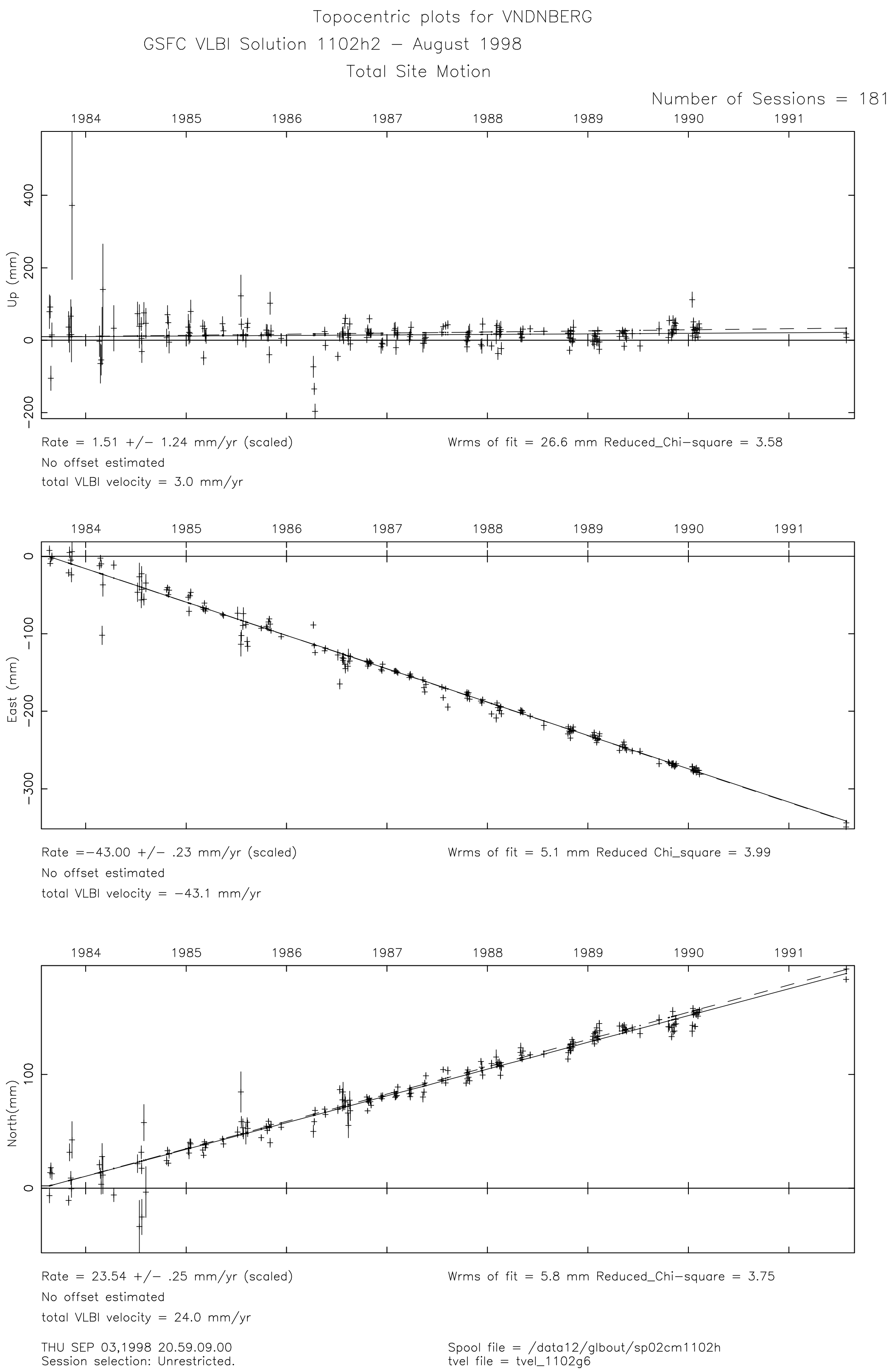


Figure 11: Station velocity components for Vandenberg (GSFC solution 1102h2).

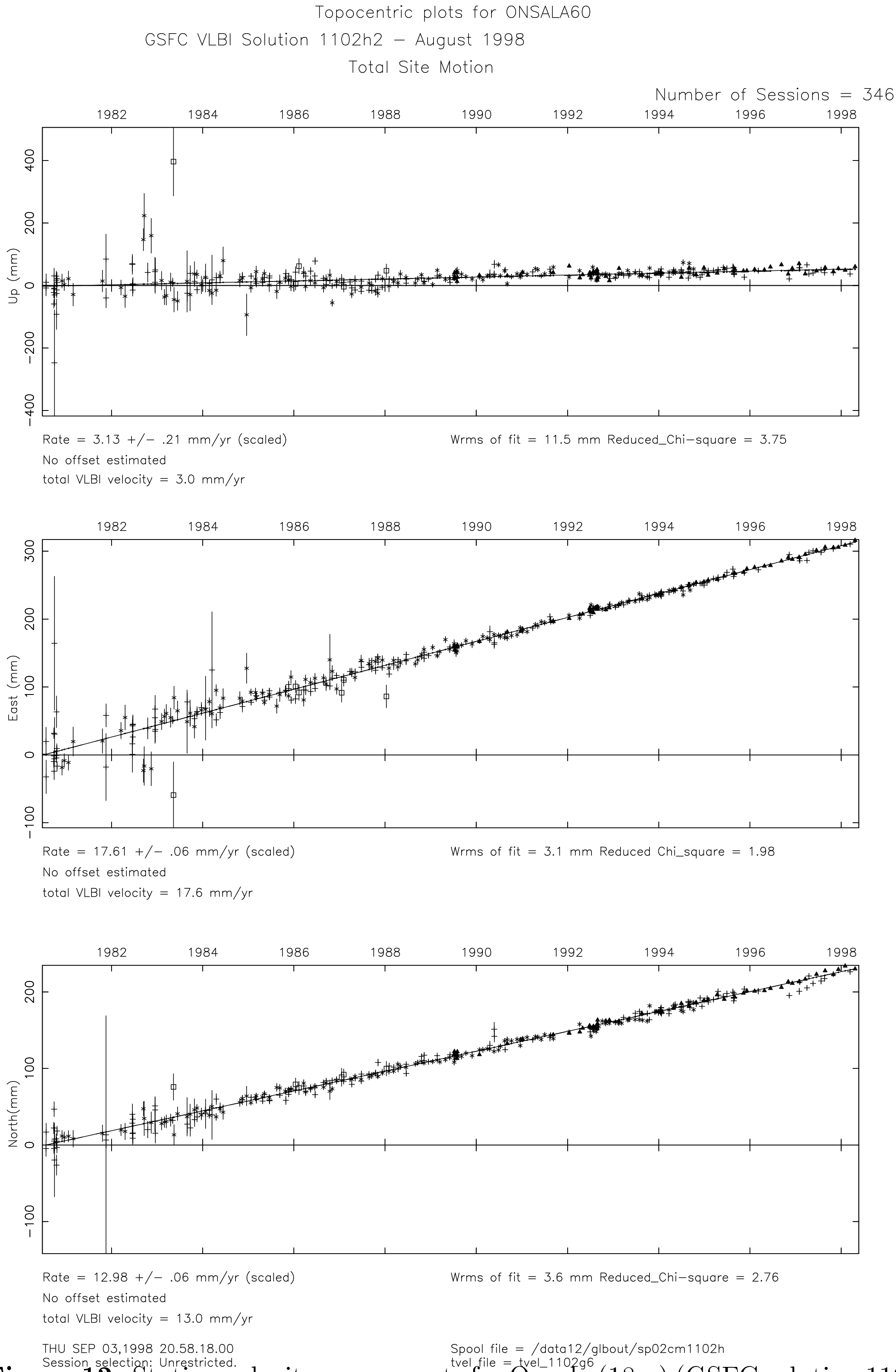


Figure 12: Station velocity components for Onsala (18m) (GSFC solution 1102h2).