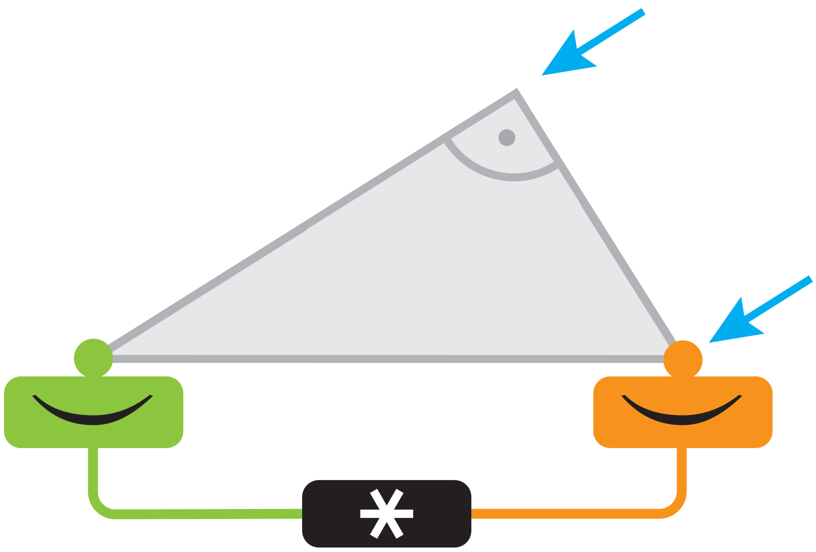
Elements of Geodetic and Astrometric Very Long Baseline Interferometry

Axel Nothnagel

TU Wien, Department of Geodesy and Geoinformation, Wiedner Hauptstraße 8, A-1040 Wien, Email: axel.nothnagel@tuwien.ac.at

Version: V4.2 - January 3, 2025



Elements of Geodetic and Astrometric Very Long Baseline Interferometry

Axel Nothnagel

TU Wien, Department of Geodesy and Geoinformation, Wiedner Hauptstraße 8,
A-1040 Wien, Email: axel.nothnagel@tuwien.ac.at

Version: V4.2 - January 3, 2025

This is an open tutorial document for educational purposes, e.g., for newcomers to geodetic and astrometric VLBI but also for the specialists wanting to expand their knowledge about topics which are not in their main focus. It is supposed to be a living document with changes and additions to be made at reasonable intervals. Everybody in the community is invited to contribute to the document either in the form of hints for additions, corrections and clarifications or with more detailed descriptions of topics. Please get in contact with me if you like to add anything or get anything added.

New major content (in addition to minor updates)

2025-01-03	Appendix H on Axis Offset determinations
2024-12-13	Small correction to "Peculiar Offsets" (10.5)
2024-12-05	Multiple small additions including 10.5 "Practical considerations of operational correlations" and "Computation of effective frequencies" in 10.5.2
2024-07-24	Multiple small additions including 2.5 "Fringe rate"
2024-04-19	Phase calibration extraction and application, *fourfit* fringe plot explanations (Appendix)
2023-10-17	New fringe fitting explanations, ambiguities, sub-ambiguities
2023-07-10	Constraints, statistics and combination issues
2023-03-04	Receiver equipment, mixed mode operations
2023-02-03	Celestial datum definition
2023-01-22	Reference frames, re-order of chapters
2022-12-02	Quick start guide, system calibrations, terrestrial datum definition
2022-02-21	VGOS and appendix on polarization
2021-11-02	Phase observations corrected
2021-09-30	Phase observations, Appendix B, Parallactic angle, IVS
2021-06-30	Earth Orientation Parameters, Partial derivatives, Phase calibration Appendix A: Rotations from xyz to Up, East, North
2021-03-09	Version 1

What is VLBI? - A quick start guide. Very Long Baseline Interferometry (VLBI) is an observing technique which has applications in geodesy, astrometry and astronomy. Observing instruments are radio telescopes which receive electromagnetic radiation emitted by compact objects in space either far distant such as quasars or nowadays also closer to Earth such as spacecraft and satellites. Since these objects emit radiation in the radio frequency domain, they are called radio sources, either natural or artificial. The natural ones are effectively infinitely far away from Earth (a few billion light years) and their radiation arrives on Earth as a plane wavefront.

The defining feature of VLBI is that pairs of radio telescopes operate together with observations having to be carried out strictly simultaneously. This also applies if multiple telescopes are employed. Geodetic VLBI aims at determining (a) telescope coordinates and their kinematic movements for establishing and maintaining terrestrial reference frames, (b) the (angular) parameters of the Earth's variable rotation, (c) the angular positions of galactic and extra-galactic radio sources establishing and maintaining celestial reference frames, and (d) the positions of artificial transmitters in space near the Earth. The field related to (c) is called astrometry. Besides geodesy and astrometry, VLBI is also a major contributor to the field of radio astronomy; in fact the origin of the early VLBI inventions lies in astronomical applications.

Geodetic and astrometric VLBI makes use of the fact that every two telescopes observing the same object should receive the same radiation pattern because it originates from a single source. Similar to Michelson's interferometry experiment, the two patterns show interferometry patterns, called fringes (oscillations of dark and bright), when the signals are superimposed while the Earth rotates. Of course, this can only be done when the analog signals are digitized and stored on some recording medium and subsequently transported to a central location either on magnetic disks or over the Internet. There, in a process called correlation, the radiation patterns are shifted relative to each other in time to find a correlation maximum which indicates that identity is found. This alone would be futile. However, when recorded in the control rooms of the two radio telescopes, the radiation patterns are marked with accurate local time stamps. For the correlation maximum, the difference of these time stamps yields the difference in arrival times of the radiation pattern at the two telescopes which is the time delay, the primary observable of geodetic and astrometric VLBI. Due to the Earth's rotation and the constantly changing geometry of the radio source and the two telescopes, the delay is constantly changing and, thus, needs an accurate time reference.

Carrying out a series of up to one-minute-long simultaneous observations of 50 - 100 radio sources by a network of several telescopes in the course of several hours (mostly 1 h or 24 h) provides the framework of a single VLBI measurement often called an observing session. Within these sessions, the process, when multiple telescopes track the same source, is called a scan, while every two telescopes forming a baseline produce a single observation. Since the radio telescopes, which often are several thousand kilometers apart, operate independently and sometimes even automatically, a detailed plan of the observation sequence, the observing schedule, has to be devised and distributed beforehand. Great care is exerted in preparing these schedules because the variations in the observation geometry through observing different regions of the sky are optimized for the best results.

From each observation, delay and phase observables are determined by correlation of the signal streams and a subsequent fringe fitting process. Compared to other space geodetic techniques, the derivation of delay and phase observables is rather complex and needs particular care. When satisfied with the quality of the observables, geodetic data analysis is carried out with special adjustment programs. Here, the geodetic and astrometric parameters of interest (telescope coordinates, Earth orientation parameters, and radio source positions) are computed from a large number of observations. For this purpose, all known geophysical effects acting on the telescopes and on the radiation on their way to Earth are modelled as accurately as possible. The functional model of geodetic and astrometric VLBI makes use of the fact that extra-galactic radio sources are almost infinitely far away so that their radiation arrives on Earth as a plane wave front. This facilitates many necessary computations and assumptions. Obtaining radio telescope coordinates at the current accuracy level of within a few millimeters has been made possible only by ongoing, decades-long research toward improvements of VLBI technology.

Abstract. This document describes in detail the theory and the individual operational steps and components needed to carry out geodetic and astrometric Very Long Baseline Interferometry (VLBI) measurements. Pairs of radio telescopes are employed to observe far distant compact radio galaxies for the determination of the differences of the arrival times at the telescopes. From multiple observations of time delays of different radio sources, geodetic parameters of interest such as telescope coordinates, Earth orientation parameters, and radio source positions are inferred. The VLBI operation's scheme generally consists of scheduling, observing session, correlation, and data analysis. Particular emphasis is given to a didactic description of technical elements and reasons why they are employed.

Contents

1. Introduction

Geodetic and astrometric Very Long Baseline Interferometry (VLBI) belongs to the group of geometric space-geodetic techniques which are used to monitor the kinematics of individual points on Earth and of the Earth as a body in space. Space geodesy has its name from the fact that observing instruments and/or their targets float in space. Primary results of space-geodetic observations are precise global reference systems realized by reference frames and parameters to describe the variable rotation of the Earth in space through so-called Earth Orientation Parameters (EOP).

To a great part, the explanations below are of course dealing with interferometry, correlation and the deduction of the geodetic and astrometric observables, which are interferometer phase, delay, and delay rate. However, to understand the coherence of the observables with the geodetic and astrometric parameters of interest, we also have to look at the framework in which the observables are deduced. To start on Earth, we have to deal with the radio telescopes forming the observing network and their coordinates. The observing network of the International VLBI Service for Geodesy and Astrometry (IVS) (Sec. 15) consists of approximately 40 active VLBI radio telescopes which are located on the surface of the Earth, however not distributed homogeneously as one can see in Fig. 1.1. As with other space geodetic techniques, the coordinates of these and all radio telescopes active in the past are referred to a conventional Cartesian terrestrial reference frame with its origin at the geocenter and the x, y, and z axes defined as a right handed system. From the origin, the x axis points at the intersection of the zero meridian and the equator, the z axis in the direction of the Earth's rotation axis, and the y axis is orthogonal to both of them. Due to the fact that continental drift and tectonic deformations produce movements of the lithosphere of several millimeters per year, the telescope positions have to be attributed with a reference date and respective velocity components v_x, v_y, and v_z in mm/year. In its entirety, the telescopes with their coordinates form a terrestrial reference frame (Sec. 13.1.1).

In the sky, we deal with compact extra-galactic radio nuclei such as quasi-stellar objects (quasars) and other galaxies emitting electromagnetic radiation in the radio frequency domain, thus radio sources. The majority of these objects are quasars and for this reason, we will generalize the naming of the natural radio sources sometimes only as quasars. These sources are observed, or more specifically tracked, with radio telescopes meaning that their radiation is received with the observing equipment to infer geodetic delay and phase observations in later processing steps. For their vast distance, most of the observed objects, as a first approximation, appear quasi-point-like and do not exhibit any proper motions on decadal time scales. Radio sources have angular positions in the sky in right ascension and declination (Sec. 4) and selections of the sources form celestial reference frames (Sec. 13.1.2).

The two frames, the terrestrial and the celestial, are linked through the Earth orientation parameters, which are time dependent angles describing the instantaneous orientation of

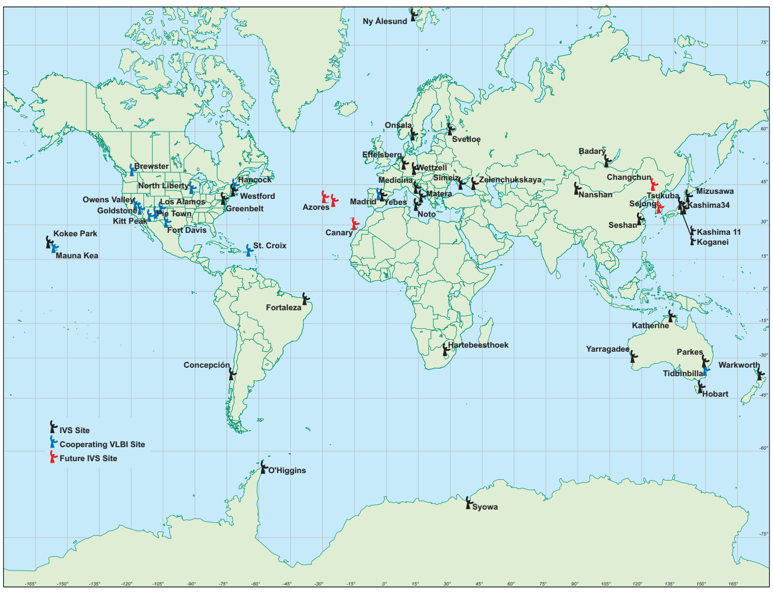


Figure 1.1: Global network of radio telescopes for geodetic and astrometric observations, see https://ivscc.gsfc.nasa.gov/stations/ns-map.html

the Earth in space. This is described in detail in analysis section 13.2. In addition to these geometric frames, we also have to look at a few realizations of time frames where differences mainly originate from relativistic transformations (Sec. 13.4).

The technique of VLBI was invented by astronomers with its first development steps in the mid 1960s for synthesizing a telescope aperture of several thousand kilometers improving the angular resolution by orders of magnitude for imaging extra-galactic objects (e.g., Matveenko et al., 1965, Broten et al., 1967, Moran et al., 1967, Bare et al., 1967). The first applications to geodesy and astrometry followed a few years later (e.g., Cohen and Shaffer, 1971, Hinteregger et al., 1972, Shapiro et al., 1974). Astrometry is an indispensable sub-discipline of both astronomy and geodesy dealing with the positions of astronomical objects.

The basic concept of geodetic and astrometric VLBI uses pairs of radio telescopes as far apart as an Earth diameter to simultaneously observe the signals of compact extra-galactic objects which emit radiation in the radio frequency regime. From the differences in the arrival times of identical radiation patterns, geodetic parameters such as telescope coordinates, positions of the celestial objects and Earth Orientation Parameters are inferred for various scientific and practical applications.

Due to the large distance, the emission of the radio sources arrives on Earth as plain wavefronts with different arrival times at the two radio telescopes forming a baseline (Fig. 1.2). The difference in arrival times is the delay τ which is one of the observables of geodetic and astrometric VLBI.

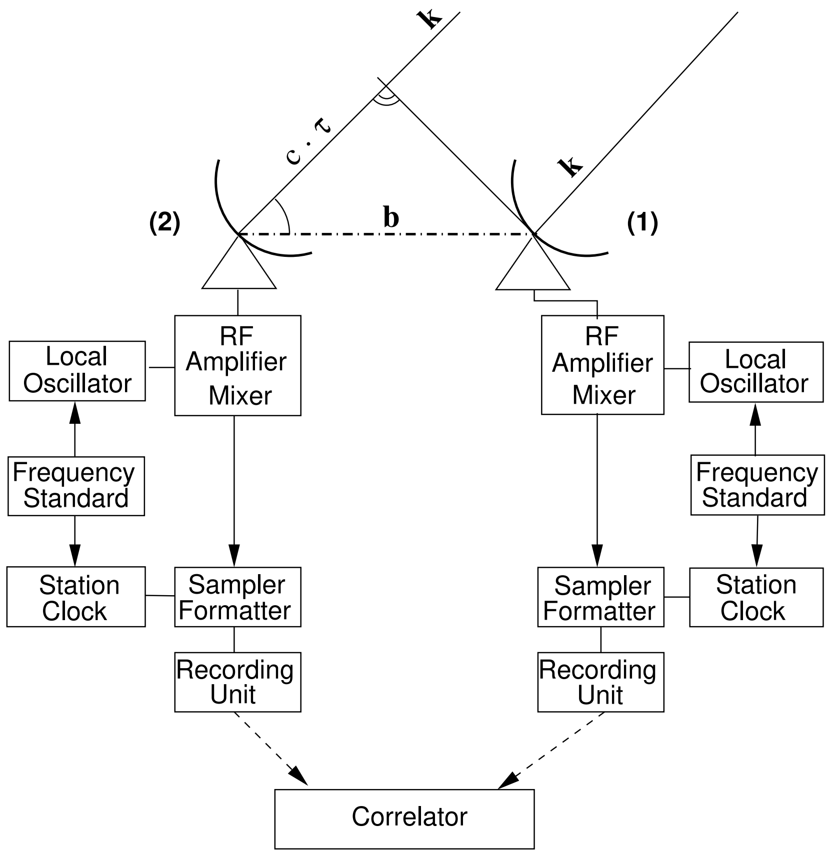


Figure 1.2: VLBI constellation with electronics components

To determine the delays, in a first step the incoming wavefronts at each telescope are digitized, time tagged, and recorded on some appropriate medium such as magnetic disks. After transportation to a correlator facility, the second step consists of a cross correlation and fringe fitting process for each individual observation to produce group and phase delay observables. The necessity for this Level-1 processing step is different to other space-geodetic techniques and adds a vast field of additional detail and complication. From a multitude of delay observables in different baseline - source geometries, the parameters of interest can then be inferred in Level-2 data analysis computations. With the extremely high accuracy and reliability of these parameters, VLBI contributes to two of the three pillars of geodesy, namely geometry and Earth rotation.

For completeness, it should also be mentioned that tracking of solar system spacecraft with VLBI has widely been used by space agencies for a few decades already. In recent years geodetic VLBI has also be performed employing man-made signal sources on artificial Earth satellites as well as on the moon. Here, the aim is to improve the orbits of these satellites by providing additional types of observations.

Since the recorded bandwidth is a key driver of the precision of VLBI observables as described below, it has always been the primary development challenge to increase the recorded bandwidth to a technical feasible and economically acceptable level. For this reason, it is realistic to say that high precision geodetic and astrometric VLBI started in 1979 with the first observing sessions of the then newly developed Mark III system (Clark et al., 1985). With a much wider bandwidth than before, precision of centimeters or even a few millimeters were reached on baselines of several thousand kilometers (Herring et al., 1986).

The technique has been engineered ever further since then, making use of growing capabilities of modern electronics, other hardware components, and analysis software including geophysical models. Efforts to exploit VLBI in a structured manner led to the foundation of the International VLBI Service for Geodesy and Astrometry (IVS) in 1999 (Nothnagel et al., 2017). Today, the IVS coordinates the majority of geodetic and astrometric developments and observations resulting in a manifold of synergies. The active network of the IVS and associated partners such as the Long Baseline Observatory (LBO, https://www.lbo.us), earlier known as Very Long Baseline Array (VLBA), consists of about 40 radio telescopes with about ten more under construction (Fig. 1.1). The tasks and structure of the IVS are described in Sec. 15.

About five years after founding the IVS, its Working Group #3[1] developed the concept for an entirely new technical realization of the geodetic and astrometric VLBI hardware and the respective adaptation of the software necessary for the new components. Under the acronym VLBI2010, the "Current and Future Requirements for Geodetic VLBI Systems" were formulated. The resulting document (https://ivscc.gsfc.nasa.gov/about/wg/wg3/IVS_WG3_report_050916.pdf) called for more agile telescopes to be able to sample the atmosphere much better than with so far existing antennas. The realization of the ideas was later called the *VLBI Global Observing System* (VGOS, see Sec. 6). Beyond faster slewing radio telescopes, the system differs from its predecessor, the S and X dual band system (also called legacy system), by a much broader bandwidth of multiple GHz. Therefore, it is sometimes also called the broadband system.

In this document, most information originates from initial considerations and theory of the VLBI concept which is applicable to both, the legacy and the broadband system. Only where necessary, a distinction is made between the two generations. In addition, I should apologize for not addressing other hard- and software realizations of the past and present such as the K line of systems of the Japanese National Institute of Information and Communications Technology (NICT) and its predecessors (Takahashi et al., 2000), the Canadian S2 System (Petrachenko et al., 2000), or the Russian developments (Matveenko, 2007; Matveenko et al., 1965; Shuygina et al., 2019).

[1] https://ivscc.gsfc.nasa.gov/technology/vgos-general.html

2. Basic VLBI theory

2.1. Interferometry

Quasars and other compact extra-galactic radio sources are the primary objects observed in VLBI. Originating from physical processes within these objects they emit pure random noise with mostly a rather flat amplitude spectrum in the frequency bands observable on Earth. Although the radiation is random noise, this is also called the signal of the radio source. Frequency bands are labeled, e.g., S band at 2-4 GHz, X band at 8-12 GHz, K band at 18-27 GHz, or Ka band at 27-40 GHz.

Applying the concept of interferometry as first devised by Michelson (1881) and Michelson (1890), the radiation received by the two telescopes is considered as two noise patterns originating from the same (point) source. They arrive on Earth with a variable phase difference due to the rotation of the Earth. The phase difference, the so-called interferometer phase is another primary observable of the VLBI technique besides the time delay. While the Earth rotates, this phase difference rotates through many cycles depending on baseline length and orientation as well as on the wavelength of the observing frequency. The determination of this phase needs some integration time applying the concepts of interferometry.

Assuming sufficient coherence for auto-correlation, the interferometer phase is determined in the correlation and the subsequent fringe fitting process to search for the interferometer phase at a certain epoch (see Sec. 10). In a rough concept of the correlation process, the digitized signal streams of the two telescopes are cross-multiplied at various trial delays for finding a correlation maximum and the respective interferometer phase.

Cycling through the phases can also be interpreted as the dark-bright variations as seen by Michelson in the optical domain initially. These variations are called *fringes* and this is the reason why in the VLBI communities we speak of "searching for fringes" if we perform the correlation and the subsequent fringe fitting process (Sec. 10). Virtually speaking, here, the correlation coefficient reaches a maximum if all parameters of the observation geometry are set exactly as during the data recording. However, the actual process works a little bit differently. The correlation process provides a single pair of amplitude and phase (called fringe visibility or just visibility) for each block of data (accumulation period) and each Fourier frequency (see Sec. 10). Over all accumulation periods and all Fourier frequencies, this virtually produces a range of mountains where we try to find the highest peak through a fitting process, thus, the expression *fringe fitting*.

Looking at geodetic VLBI from a different perspective, we start with knowing that compact extra-galactic objects are found as far away as the frontiers of the universe at quasi-infinite distances. For this reason, the emission of these radio sources arrives on Earth as plain wavefronts and causes different arrival times at the two radio telescopes forming a baseline (Fig. 1.2).

The difference in arrival times can be expressed as a phase delay τ_{ph} (Sec. 2.2). Its time derivative is the phase delay rate $\dot{\tau}_{ph}$ (Sec. 2.3). Quasars and other compact radio sources do not emit just monochromatic radiation but a broad spectrum of frequencies. If the delay is determined from a spectrum of superimposed frequencies, we speak of the group delay τ_{gr} which is a time delay (Sec. 2.4). These three quantities are the fundamental observables of geodetic and astrometric VLBI.

2.2. Phase delay

The initial quantities, which are related to the output of interferometry, are the time dependent phases of the cross-power spectra across the visibility frequencies (Sec. 2.1). From these a single phase with respect to a reference frequency at a reference epoch is extracted (Sec. 11.6. This quantity reflects what in general terms is called the interferometer phase. At this stage, it should be sufficient to explain a few peculiarities of the phase observable.

If quasars would emit just a single monochromatic wave at a certain frequency, we can imagine that the $c \cdot \tau$ vector pointing towards telescope (2) in Fig. 1.2 is superimposed with a continuous sine wave having a wavelength $\lambda = c/\nu$ (c = speed of light, ν = frequency of the wave). The angular frequency ω in radian [-] relates to the observing frequency ν in Hertz [Hz] by

$$\omega = 2\pi\nu = \frac{2\pi}{T} \tag{2.1}$$

where T is time elapsed.

Since this sinusoid of the phase is in fact a difference, the phase angle at the fictitious emission point (where the square angle is in the graph) is zero. Then the phase determined at telescope (2) is a phase angle with 2π ambiguities. Although the ambiguities can be as large as 10^7, the phase observations are still of value, in particular for short local baselines and in general in astronomical applications such as source mapping.

Phase observations come along as total phase in radian or phase delays in seconds (Sec. 11.6). Both quantities map the total length of the $c \cdot \tau$ vector because the respective a priori quantity is added at fringe fitting output.

Another way to consider the phase and delay relationship with ν_0 being the reference frequency is that

$$\phi_0 = 2\pi\nu_0 \cdot \tau \tag{2.2}$$

From this, it follows that

$$\tau_{ph} = \frac{\phi_0}{2\pi\nu_0} = \frac{\phi_0}{\omega_0} \tag{2.3}$$

The most important fact, why phase delays and total phases are of interest to geodetic and astrometric VLBI analyses, is that the formal error of the phase observations is a factor of about 30 better than that of legacy X band group delay observations (Eq. 3.6 in Sec. 3). However, for baselines longer than a few tens of kilometers, the ionosphere enters the game and phase ambiguity resolution becomes a critical issue.

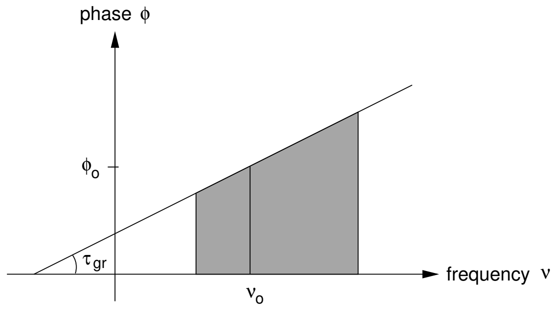


Figure 2.1: Phase versus frequency according to Petrov (1999)

2.3. Phase delay rate

The phase delay rate is defined as the slope of the phase ϕ at a reference frequency ν_0 (or angular frequency ω_0) with respect to time t:

$$\dot{\tau}_{ph} = \frac{1}{\omega_0}\frac{d\phi(\omega)}{dt} = \frac{1}{2\pi\nu_0}\frac{d\phi(\nu)}{dt} \tag{2.4}$$

The phase delay rate is unitless but mostly expressed in seconds per second (s/s). This is the quantity which is reported in the fringe fitting output and used in the geodetic analyses. For this reason, the index *ph* will be omitted later on in this tutorial. The need for and the benefits of the phase delay rate are also an immediate consequence of the fact that the phases of the cross-power spectrum are fundamental to correlation and fringe fitting.

In contrast to the definition in Eq. 2.4, we could also make use of the group delay rate $\dot{\tau}_{gr}$ which is the derivative of the group delay τ_{gr} (Sec. 2.4) with respect to time:

$$\dot{\tau}_{gr} = \frac{d\tau}{dt}. \tag{2.5}$$

However, this would require the determination of the group delay at more than one epoch within an observation and would not generate any benefit with respect to the phase delay rate.

2.4. Group delay

The group delay τ_{gr} is by far the most important observable in analyses of geodetic and astrometric VLBI observing sessions and is, thus, just designated τ throughout this document unless marked otherwise. Since there is only one delay rate, this is generally just called the delay rate $\dot{\tau}$.

In the context of bandwidth synthesis, other expressions for the group delay exist which specify its derivation such as single-band delay or multi-band delay. The details are described in Sec. 3.

The group delay is defined as the slope

$$\tau = \frac{d\phi(\omega)}{d\omega} = \frac{1}{2\pi}\frac{d\phi(\nu)}{d\nu} \tag{2.6}$$

with ϕ being the phase, ω being the angular frequency, and ν being discrete observing frequencies (Fig. 2.1).

For its importance, the group delay deserves a closer look because it comes along in different geometric constellations. Turning to the observations, we can derive the delay τ as a difference in arrival times $T^{(1)}$ and $T^{(2)}$ at the two telescopes (1) and (2). The principal connection between the time delay τ, the baseline $\mathbf{b}$ and the unit vector in source direction $\mathbf{k}$ is

$$\tau = T^{(2)} - T^{(1)} = -\frac{1}{c}\mathbf{b}\cdot\mathbf{k} \tag{2.7}$$

with c being the speed of light. This is a simplification and needs some further specifications to be usable in a proper way in later data analyses. The first one is that, of course, the Earth is rotating and that the wavefront arriving from the quasar travels with finite velocity. This leads to the fact, that if the wave front arrives at telescope (1) (Fig. 1.2) at time $T^{(1)}$ and continues travelling to telescope (2), the latter has moved from its position at $T^{(1)}$. At the same time, the Earth has moved on its orbit around the Sun causing another effect of displacement. The first effect is the so-called daily aberration, the second is annual aberration. Sometimes these two effects are also called retarded baseline effect. For these reasons, the arrival time $T^{(2)}$ at telescope (2) refers to an entirely different geometry and Eq. 2.7 will never represent the complete situation.

The second issue with the delay (and of course the other observables as well), which is a consequence of the moving Earth, is the need for a reference epoch at which the delay is valid. In a station-based concept, the reference epoch is that when the (imaginary) wavefront passes telescope (1) (Fig. 2.2). However, as we will see later, VLBI needs some finite integration time and the Earth of course rotates during that time. The reference epoch has, thus, to be defined within this integration time. In a triangle or in a network of telescopes, there may then be multiple natural reference epochs depending where the signals have arrived when. However, in the fringe fitting process, there may be a logic which forces the reference epochs of all observations of an entire network observing the same radio source, a so-called *scan*, having identical reference epochs for all observables.

This station-based delay, or baseline delay, as well as the respective delay rate and phase delay, are the natural observables in most geodetic and astrometric data analysis programs. In these software packages, the aberration effects are applied to compensate for the aberration distortions (see Eq. 13.45).

In contrast to the station-based reference epoch, there is a natural reference epoch, which is unique for all baselines of a network, and this is the epoch when the wavefront passes the geocenter. Initially, the geocenter delay is a delay for a single telescope with respect to an imaginary telescope at the geocenter (GC), $\tau_{(1)-GC}(t_{GC})$ (Fig. 2.3). In fact, this can be considered as a baseline between the geocenter and the radio telescope, e.g., GC-(1). For the purpose of getting a priori observables, this computation, including the aberration corrections, can easily be done for all other telescopes in the triangle or network as well and thus provides initial geocenter delays for all telescopes at the same reference epoch. The actual geocenter baseline delays can then be formed by computing

$$\tau_{i-j}(t_{GC}) = \tau_{i-GC}(t_{GC}) - \tau_{j-GC}(t_{GC}) \tag{2.8}$$

which can also be represented graphically (Fig. 2.3). While the initial delays ($\tau_{i-GC}(t_{GC})$) are always positive for obvious reasons, the difference in Eq. 2.8 depends on the order of the telescopes. These are normally ordered according to the alphabet.

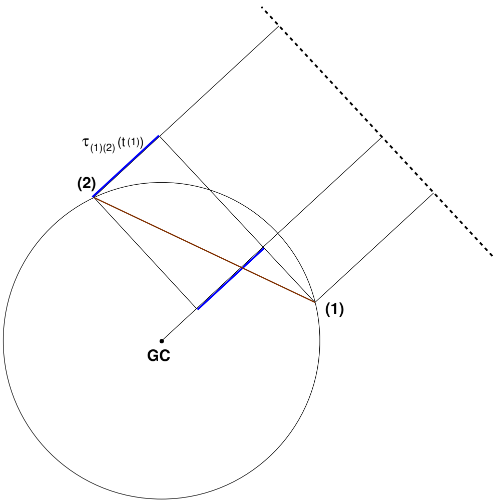


Figure 2.2: Baseline delay at time of arrival of wave front at telescope (1) (blue line).

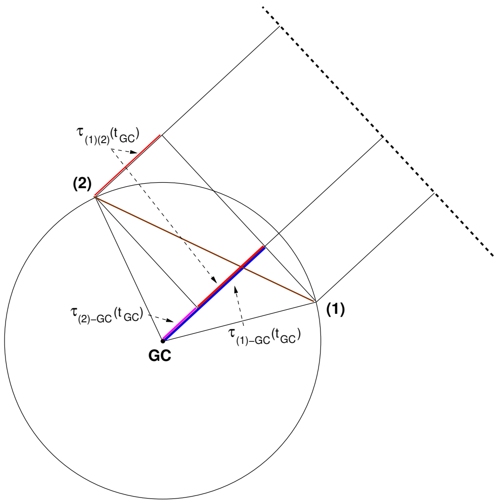


Figure 2.3: Geocenter delay at time of arrival of wave front at the geocenter (GC) (red line).

2.5. Fringe rate

In the context of VLBI, often the expression *fringe rate* or *fringe frequency* is used. It is nothing more than the rate with which the interferometer phase repeats its cycles. In standard VLBI observations of quasars with ground based radio telescopes, interferometer phase oscillations to first order and neglecting instrumental drifts purely originate from the rotation of the Earth. Talking about the interferometer phase, the fringe rate or fringe frequency always refers to correlator output which is always a superposition of two signals. The fringe rate or fringe frequency ν_f is

$$\nu_f = \nu_0 \frac{d\tau}{dt} = \nu_0 \cdot \dot{\tau}. \tag{2.9}$$

where ν_0 is the reference frequency (Thompson et al., 2017).

A simplistic geometric derivation of $d\tau/dt$ starts with a 2D case of Eq. 2.7 (see also Appendix G), applying the angle θ between baseline and unit vector in source direction and the baseline's East-West component b_{EW}, reading

$$\tau = -\frac{1}{c} \cdot b_{EW} \cdot \cos\theta. \tag{2.10}$$

Then, we make use of sequential derivatives

$$\frac{d\tau}{dt} = \frac{d\tau}{d\theta} \cdot \frac{d\theta}{dt} \tag{2.11}$$

where

$$\frac{d\tau}{d\theta} = \frac{1}{c} \cdot b_{EW} \cdot \sin\theta \tag{2.12}$$

and $d\theta/dt = \omega_E = 7.29115 \times 10^{-5}$ rad/s is the rotation velocity of the Earth. Introducing this into Eq. 2.9 and applying the cosine of the radio source's declination ($\cos\delta$) to account for non-equatorial sources, we find

$$\nu_f = \omega_E \nu_0 \frac{1}{c} b_{EW} \sin\theta \cdot \cos\delta \tag{2.13}$$

For $\nu_0 = 8.4$ GHz, $b_{EW} = 6000$ km, $\theta = 90°$ and $\delta = 0°$, we find a ν_f of 12.2 kHz. This fringe rate is also called the natural fringe rate or fringe frequency. We will encounter this again in the fringe stopping process during correlation (Sec. 10.4).

Sometimes, astronomers talk about baseline lengths in wavelengths. For this, they compute $n_\lambda = \nu_0 \frac{1}{c} b$ yielding a 100 million wavelength baseline if it is 3400 km long and the X band wavelength is 0.034 m.

It should be mentioned that the natural fringe rate and with it also the delay rate is insensitive to the North-South component of the baseline. This is different if moving objects such as Earth satellites are observed. Then a full 3D derivation has to be considered.

3. Precision considerations of VLBI observations

Most of the technical equipment is designed to produce observables with the best accuracy possible. Fundamental to the following is that the radiation observed and recorded by a VLBI system consists of a "signal" and a "noise" component. This may be confusing at first sight but can easily be explained by considering the noise, which is emitted solely by the radio source, to be the "signal". All other noise added along the way then is the real "noise". Both components have Gaussian statistics (Whitney, 1974).

The relationship between "signal" and "noise" is mapped through the signal to noise ratio (SNR). It is a dimensionless quantity in VLBI. In the mathematical definition of Whitney (1974), the SNR is a product of the normalized correlation amplitude ρ_0, the spanned bandwidth $\Delta\nu$ and the integration time T

$$SNR \equiv \rho_0 \sqrt{2\Delta\nu T} \tag{3.1}$$

The number under the square root represents the number of bits entering the correlation process, often called total number of samples. The normalized correlation amplitude depends on the thermal noise-equivalent antenna temperatures $T_{A_{(i)}}$ and the system temperatures $T_{S_{(i)}}$ of the two telescopes ($i = 1, 2$)

$$\rho_0 \approx \sqrt{\frac{T_{A_{(1)}} \cdot T_{A_{(2)}}}{T_{S_{(1)}} \cdot T_{S_{(2)}}}} \tag{3.2}$$

The noise level received at the antenna is represented as the equivalent of the respective temperature of thermal noise called antenna temperature. It depends on the flux density S of the radio source being observed. Since we encounter some loss due to digitizing the analog signal, the final theoretical SNR can be computed from

$$SNR = \eta \frac{S}{2k} \sqrt{\frac{A_{(1)} \cdot A_{(2)}}{T_{sys_{(1)}} \cdot T_{sys_{(2)}}}} \sqrt{2\Delta\nu T} \tag{3.3}$$

with η = digitizing loss factor, which is 0.57 for 1-bit, 0.64 for 2-bit sampling (Thompson et al., 2017)), S = correlated flux density of the radio source, k = Boltzmann's constant (1.38 x 10^{-23} Ws/K), A_i = effective antenna areas of telescope (1) and (2), T_{sys} = noise temperatures of the receiving systems, $\Delta\nu$ = total bandwidth of the receiving system, T = coherent integration time.

The SNR is the driving factor of all considerations related to the precision of the VLBI observables. Since the variance of a noise-corrupted fringe phase estimate was found to be

$$\sigma_\phi^2 \approx \frac{T_{S_{(1)}} \cdot T_{S_{(2)}}}{T_{A_{(1)}} \cdot T_{A_{(2)}}} \cdot \frac{1}{2\Delta\nu T} \tag{3.4}$$

it follows that

$$\sigma_{\phi} \approx \frac{1}{SNR} \tag{3.5}$$

The uncertainty of the phase delay is even more accurate because the phase is divided by the frequency (Eq. 2.3) and thus

$$\sigma_{\tau_{ph}} \approx \frac{1}{2\pi\,\nu_0 \cdot SNR} = \frac{1}{\omega_0 \cdot SNR} \tag{3.6}$$

The uncertainty of the phase delay rate $\dot{\tau}$ depends, besides on the SNR, on the rms integration time ΔT_{rms} with

$$\sigma_{\dot{\tau}} \approx \frac{1}{2\pi\,\nu_0 \cdot \Delta T_{rms} SNR} = \frac{1}{\omega_0 \cdot \Delta T_{rms} \cdot SNR} \tag{3.7}$$

where ΔT_{rms} is

$$\Delta T_{rms} = \sqrt{\frac{1}{T}\sum(t_\ell - \bar{t})^2} \tag{3.8}$$

with t_ℓ being the sample epochs, $\bar{t}$ being the mean epoch (Whitney, 1974). For continuous integration periods over T seconds

$$\Delta T_{rms} = \sqrt{\frac{T^2}{12}} \tag{3.9}$$

and thus

$$\sigma_{\dot{\tau}} \approx \frac{12}{2\pi\,\nu_0 \cdot T \cdot SNR} = \frac{12}{\omega_0 \cdot T \cdot SNR} \tag{3.10}$$

Finally, we get to the uncertainty of the group delay. Here, besides of the SNR, the total spanned bandwidth is the driving factor determining the theoretical accuracy of the group delay (Whitney, 1974):

$$\sigma_{\tau} = \frac{1}{2\pi \cdot SNR \cdot \Delta\nu_{RMS}} \tag{3.11}$$

with $\Delta\nu_{RMS}$ = effective bandwidth of receiving system depending on the number of sub-bands n, the individual sub-band frequencies ν_i and the mean frequency $\bar{\nu}$

$$\Delta\nu_{RMS} = \sqrt{\frac{\sum_{i=1}^{n}(\nu_i - \bar{\nu})^2}{n}}. \tag{3.12}$$

All these quantities heavily depend on the SNR. Although the actual SNR is computed during fringe fitting, Eq. 3.3 helps us to determine where we can improve the system and the layout of the observations in the scheduling process for better results. First, it should be noted that the SNR formulation is independent of the observing frequency. However, for the system temperature T_{sys}, the efficiency of the radio telescopes for VLBI observations needs to be determined regularly. The efficiency is expressed as System Equivalent Flux Density

(SEFD), S_{sys}. The SEFD represents the flux density of a fictitious radio source which would double the output power of the complete receiving system. The spectral flux density of a radio source as well as for a receiving system is given in Jansky [Jy] where 1 Jy = 1 x 10^{-26} W/(Hz * m^2). S_{sys} is computed according to Thompson et al. (2007) with

$$S_{sys} = \text{SEFD} = \frac{2k}{A_{eff}} T_{sys} \tag{3.13}$$

where A_{eff} = effective antenna aperture, T_{sys} = system temperature , which is composed of the sky brightness temperature T_{sky} and the noise temperature of the instrument T_{inst},

$$T_{sys} = T_{sky} + T_{inst} \tag{3.14}$$

with $T_{sky} = T_{so} \cdot \lambda^{2.55}$ and T_{so} being the galactic background radiation of 60 K ± 20 K. λ is the wavelength of the radiation.

The SEFD has the advantage that it can be measured easily at each telescope independently by determining the fractional increase in power obtained when pointing on and off a source of known flux density. SEFDs range from 60 Jy to 400 Jy for large antennas, 800 Jy to 1000 Jy for the 20 m class systems to more than 15000 Jy for telescopes of only a few meters of diameter. These values also strongly depend on the receiver noise mitigation by cryogenic systems.

With the SEFDs at hand, e.g., tabulated in SKED catalog files, we can also formulate the projected SNR by entering Eq. 3.13 into Eq. 3.3 yielding

$$SNR = \eta\, S\, \frac{1}{\sqrt{\text{SEFD}_{(1)} \cdot \text{SEFD}_{(2)}}} \sqrt{2\Delta\nu T} \tag{3.15}$$

The actual SNR of an observation is determined at the fringe fitting process. Fundamental to the SNR determination is the amplitude of the correlation function (Sec. 10.3). In basic terms, the SNR is the ratio between the amplitude of the correlation peak and the mean amplitude of the correlation function without that peak.

The SNR threshold for a successful detection of fringes is approximately 10 (Whitney, 1974). However, the technical realization above, the selection of bright radio sources, and the duration of the observation is generally set to achieve a minimum SNR of 25 or 35. Sometimes even with minimum integration time, the SNR may exceed 50 or 100 for large aperture telescopes and the brightest radio sources.

Returning to further options for reduction in standard deviation of the delay observables, we look again at Eq. 3.3. With most of the parameters being predefined, the integration time and the bandwidths are the parameters which still allow some scope for increase. The first parameter of these two, the integration time, should be kept short. An initial reason was that the frequency standards allowed only for a limited coherence time but this has been overcome by the use of hydrogen maser clocks (H-Maser). Today, short integration times permit collecting observations at as many different spatial directions as possible in a short time frame for a good sampling of the atmospheres above the radio telescopes (Petrachenko et al., 2009). And, thus, this is one of the targets of today's observing schedules.

The second, which can still be optimized, is the observed bandwidth, which appears as the effective bandwidth in Eq. 3.11 and as total bandwidth in Eq. 3.3. At the time of the developments of the legacy S/X systems, recording capacity had been (and still is) a limiting factor. For this reason, bandwidth synthesis (BWS) has been invented (Rogers, 1970) and is still in use today also in the VLBI Global Observing System (VGOS). As we will see in Sec. 6, the developments of the VGOS system are also limited by storage capacity and make use of BWS.

In order to understand the basics of and the necessity for bandwidth synthesis we should first look at the single band delay determination. Consider that we have recorded a single channel of a certain bandwidth B with a rectangular bandpass. Without going into the details here, the shape of the cross correlation function, which is the result of the Fourier transform of the bandpass from the frequency domain into the lag domain, has the form of a so-called sinc-function ((sin x)/x) with a delay width Δ_τ which follows

$$\Delta_\tau = \frac{1}{B} \tag{3.16}$$

with B being the bandwidth of the bandpass. The delay width is important since, besides the SNR, it determines the inherent accuracy of the delay determination. The reason is that the location of the peak of the sinc-function determines the value of the delay on the delay axis. And, of course, the sharper the peak, the more accurate the determination of the location of the maximum on the τ axis (Fig. 3.1).

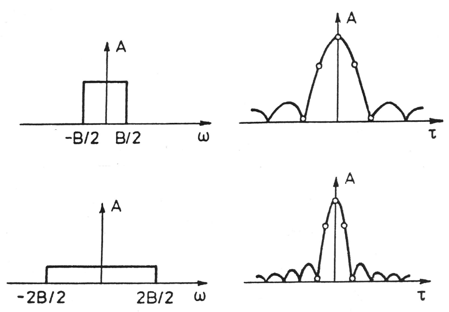


Figure 3.1: Fourier transform from frequency domain (left) into lag domain (right) with bandwidth B and 2B (Beyer, 1985)

Consequently, as can be identified from Fig. 3.1, the peak is the sharper, the broader the bandpass. Naturally, one could now argue that we just need to increase the bandwidth to one or more GHz and can drive the precision of the delay determination down to a picosecond. However, limited recording capabilities have always inhibited the technical realization of this idea.

A suitable way to bypass the storage problem at least to some extend is bandwidth synthesis (BWS). The background of BWS is described in Sec. 11.2. For the moment, it is just enough to understand the following. The BWS technique in legacy S/X mode works on the basis of extracting and processing comparatively small frequency sub-bands (2, 4, 8 or 16 MHz depending on equipment capabilities) out of a spectrum of, e.g., 720 MHz at X band and 125 MHz at S band. The VGOS system works similarly, however with different sub-band allocations (Sec. 6.2). With this trick, a large total bandwidth is spanned but reducing the need for storage and transportation to only n times m MHz (Fig. 3.2). Of course, maximizing the total bandwidth within its technical limits is still a target.

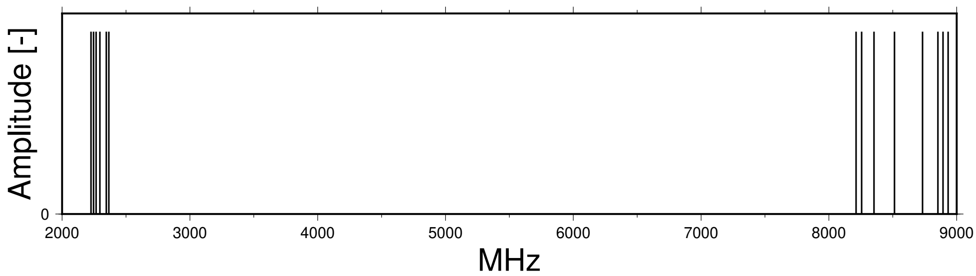


Figure 3.2: Sub-band allocation of S and X band with 6 and 8 sub-bands, respectively, for bandwidth synthesis, sub-band bandwidth 8 MHz. N.B.: S and X band are always handled separately in correlation, fringe fitting and analysis.

While the two extreme frequency bands help to increase the effective bandwidth $\Delta\nu_{RMS}$ according to Eq. 3.12, the interposed bands support this as well but also increase the total bandwidth $\Delta\nu$ in Eq. 3.3. As shown later, in the legacy S/X systems, the incoherent sum of individual sub-bands of either S or X band produces so-called single-band (group) delays while the BWS combination of all sub-bands in either S or X band yield multi-band (group) delays.

At this point, we should briefly explain how frequency sub-bands are labeled. For this purpose, regularly, the nominal frequency of a channel/sub-band is reported together with the attribute of its sideband, which is either lower or upper sideband (LSB/USB). Each sub-band has a certain bandwidth which may be 4, 8, 16 or even 32 MHz. A frequency sub-band of 2225.99 - 2233.99 MHz is thus often just denoted 2225.99 GHz while the "upper sideband" is omitted because it is the default in legacy S and X band operations .

4. Extra-galactic radio sources

The targets of geodetic and astrometric VLBI are compact extra-galactic objects. In general, these objects are galaxies which, from an astronomical point of view, appear in a variety of different kinds. Since they are radio-loud with high power, quasars (quasi stellar objects, QSOs) are observed most commonly in geodetic applications. Quasars have an active galactic nucleus, which may be a black hole, surrounded by an accretion disk of rotating accumulated gas and matter (Fig. 4.1). Due to gravitational processes, the quasar ejects two relativistic jets perpendicular to the accretion plane. With respect to Earth, the jets point to us in an arbitrary position angle.

Figure 4.1: Artist's rendering of quasar P172+18, Credit: ESO/M. Kornmesser

What we can observe is the electro-magnetic emission of regions of the jet which points towards Earth. The counterpart is blanked out by Doppler boosting. In the early days of VLBI, it was said that, due to large distances, the observed emission and thus the objects appear quasi-point-like and do not exhibit any proper motions. However, with increased resolution of the interferometers, the morphology of the radio sources and also its time and frequency dependence become visible. As a consequence, it is known today that the cores of emission, which we observe, are the closer to the nucleus the higher the observing frequency. This leads to an effect, which is also known as core shift. The exact magnitude depends on the source and is at the level of single digit microarcseconds.

In addition, the asymmetric field of radiated energy leads to the so-called source structure effect which causes a deviation from the appearance as a point source. The smaller the position angle between the jet direction and the line of sight from Earth, the higher the probability that we find a point source. Extended source structure corrupts the phase of the signal and consequently the inferred delays. For this reason, Charlot (1990) introduced a categorization by source structure index distinguishing sources with no structure from those with extended structure (Fig. 4.2). Currently, objects with an index number of four are excluded from geodetic observations whenever possible but endeavors also exist to correct for these effects with adequate models.

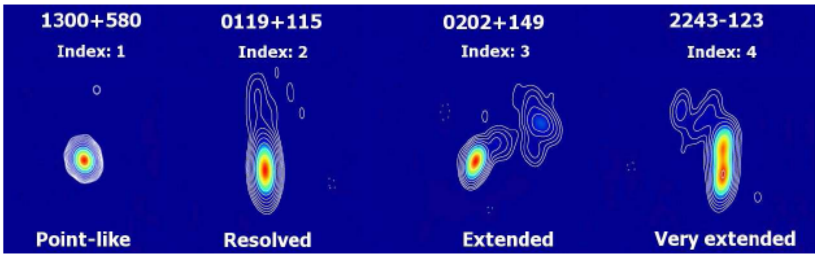


Figure 4.2: Images of radio sources with different structure indices (Collioud and Charlot, 2009).

The positions of the objects are defined as angular right ascension α and declination δ components of an equatorial polar coordinate system. Right ascension (R.A.) is counted positive in the direction of the Earth's spin in the equatorial plane from 0 to 24 hours with minutes and seconds of time (Fig. 4.3). Declination runs in degrees, minutes, and seconds of arc from -90° at the celestial south pole to +90° at the celestial north pole.

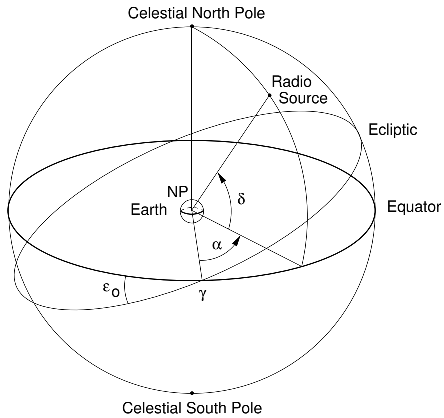


Figure 4.3: Coordinates α and δ of a radio source on the celestial sphere. NP is the terrestrial North Pole. ε_0 is the mean obliquity of the ecliptic of about 23.5° and γ is the vernal equinox.

Many compact extra-galactic radio sources are named with their celestial positions in the form HHMM±DDd where HH and MM are the hours and minutes of right ascension while DDd represents the declination in degrees (DD) with the first decimal (d). For historical reasons they are often positions referred to the epoch B1950.0. This applies to naming conventions of the IVS and IERS. The International Astronomical Union (IAU) uses names referred to epoch J2000.0. For unambiguous use of source names, the IVS maintains a translation table[2], where, e.g., 0018-194 is the IVS name, J002109.3-191021 is the long IAU name, and J0021-1910 is the short IAU name, following similar naming conventions in terms of embedded position.

[2]https://cddis.nasa.gov/archive/vlbi/gsfc/ancillary/solve_apriori/IVS_SrcNamesTable.txt

5. Observing instruments

5.1. Radio telescopes

5.1.1. Telescope optics and mounts

Radio telescopes for geodetic and astrometric applications are basically the same as those for pure astronomical observations. The only differences are found in the receiver and backend technology. The dominant part of a radio telescope is the main reflector, often combined with a secondary reflector or sub-reflector, which serves to combine the energy of the incoming wave front in a single focal point. The construction of a radio telescope aims at locating the phase center of a feed horn exactly at the focal point to guide the energy into the receiver which predominantly consists of the first stage of amplification in a cooled (cryogenic) environment.

A radio telescope is said to have a prime focus system if the radio frequency feed horn is placed at the focal point of the main reflector. Most radio telescopes, however, use a secondary focus near the vertex of the paraboloid for easier access to the receiver equipment. For this purpose either a hyperbolic (Cassegrain system, Fig. 5.1) or an ellipsoidal sub-reflector (Gregorian system, Fig. 5.2) are mounted in front of or behind the primary focal point, respectively, to concentrate the energy in a secondary focal point. The feed horn is then placed here.

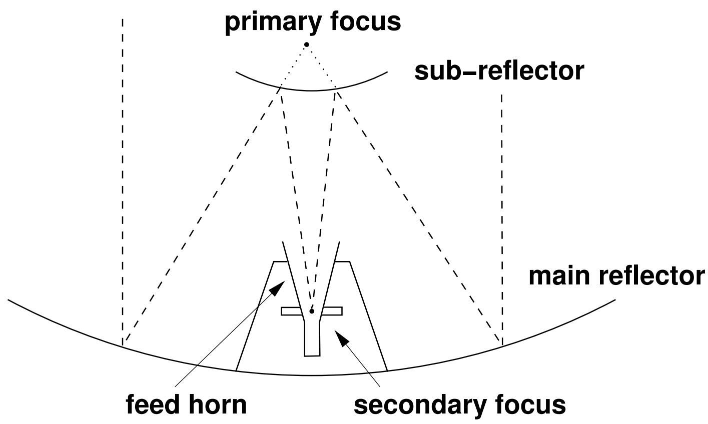


Figure 5.1: Cassegrain radio telescope optics with parabolic main reflector and hyperbolic sub-reflector (dashed lines are discrete ray paths)

For completeness, it should be mentioned that there are secondary focus telescopes which employ additional wave guide optics underneath the main reflector. They are called Nasmyth radio telescopes (Baars, 2007). The idea here is to remove the receiver(s) from the movable telescope parts to a stable platform where tests and maintenance can be performed more conveniently.

The reflector optics is mounted on a pair of axes which are perpendicular to each other to reach all positions on the sky. The most common mount is the azimuth-elevation, also

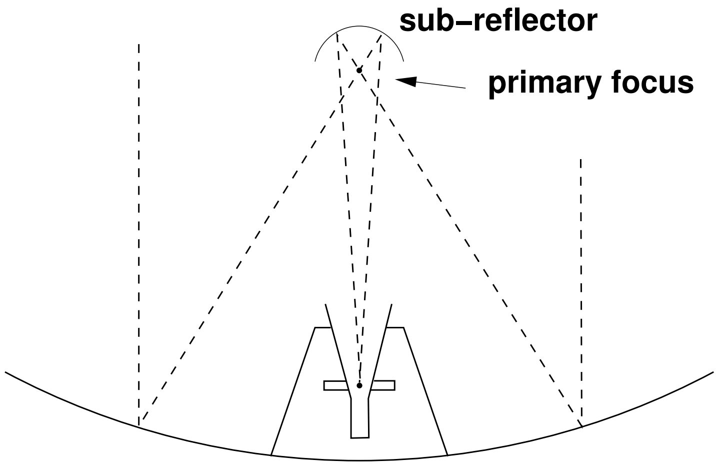


Figure 5.2: Gregorian radio telescope optics with parabolic main reflector and ellipsoidal sub-reflector (dashed lines are discrete ray paths)

called alt-azimuth, axis system (Fig. 5.3) (Baars, 2007). Other systems are polar mounts, where the primary axis is in a position parallel to the Earth's rotation axis (Fig. 5.4), and the X/Y mount, where the primary axis just lies in a horizontal plane (Fig. 5.5).

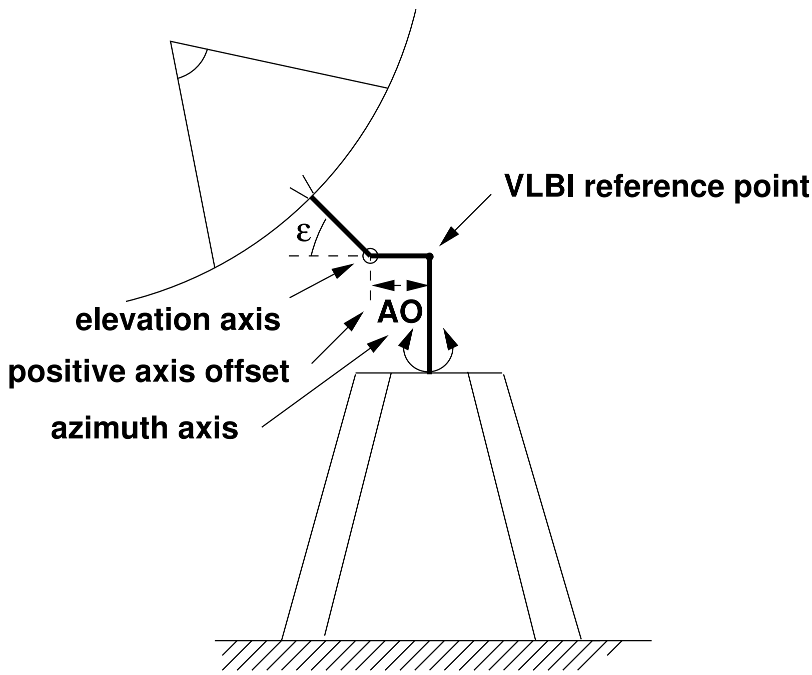


Figure 5.3: Azimuth-elevation mount with primary axis in local vertical.

It should be mentioned here that azimuth-elevation telescopes are constructed in two different ways. The one group is a so-called *wheel and track* type where the whole structure moves on a circular track with at least four groups of wheels distributing the weight evenly (Fig. 5.6). Today, this construction is mainly used for bigger telescopes of diameters larger than 20 m. As with telescopes with polar and X/Y mounts these are generally made of steel

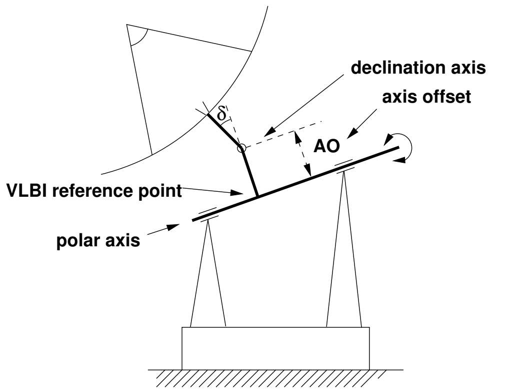


Figure 5.4: Polar mount with primary axis parallel to Earth rotation axis.

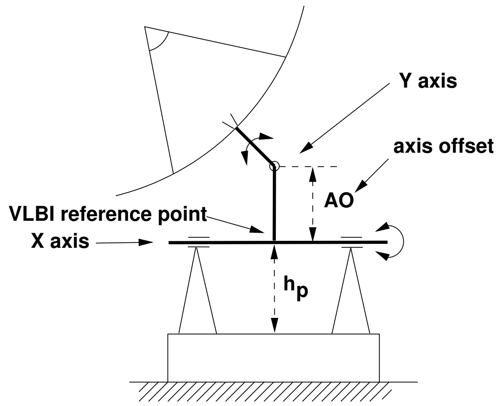


Figure 5.5: XY mount with primary axis in horizontal plane.

entirely. The second group consists of so-called *turning-head* telescopes where only the top part rotates while a tower made of concrete supports the moving parts (Fig. 5.7).

Figure 5.6: Wheel and track telescope (Seshan 25 m, P.R. China).

Figure 5.7: Turning head telescope (Wettzell 20 m, Germany).

In geodetic and astrometric VLBI, the Earth-fixed coordinates of a radio telescope are always related to the VLBI reference point. This is a point within the structure of a telescope which is invariant to any rotations of the telescope (Sovers et al., 1998). To first order this is the intersection of the primary and the secondary rotation axis of the telescope. However, since these intersect only in very seldom cases, the reference point is the projection point of the secondary axis onto the primary axis (Fig. 5.3, 5.4, 5.5) The separation of the two axes is called the axis offset (*AO*). It can range from a few millimeters in cases, where intersecting axes were planned but not exactly realized for constructional reasons, to several meters where required for technical reasons, in particular for polar and X/Y mounts. In seldom cases, the axis offset may even be negative resulting from the fact that the elevation axis lies a small distance behind the azimuth axis (Nothnagel, 2009).

Considering the travel time of the incoming wave fronts, the distance between the feed horn and the secondary axis is considered invariant to first order. It should be mentioned here already that any constant part of the signal path produces a constant time delay contribution which is treated as relative clock offset in the analysis (see Sec. 13). The axis offset, however, produces a time delay $\Delta\tau_{AO}$ which varies with elevation. In general, it depends on the unit vector in source direction **s** and the unit vector in the direction of the fixed axis **f** (Nothnagel, 2009):

$$\Delta\tau_{AO} = \frac{1}{c} AO \cdot \sqrt{1-(\mathbf{s}\cdot\mathbf{f})^2}. \tag{5.1}$$

For azimuth-elevation telescopes the axis offset produces an extra time delay contribution of

$$\Delta\tau_{AO} = \frac{1}{c} \cdot AO \cdot \cos\varepsilon. \tag{5.2}$$

with ε being the elevation angle of the pointing, while for polar mounts the delay contribution depends on the declination of the radio source δ:

$$\Delta\tau_{AO} = \frac{1}{c} \cdot AO \cdot \cos\delta. \tag{5.3}$$

For X/Y mounts one has to distinguish in which direction the primary axis is oriented. For those in north-south direction, the delay contribution is

$$\Delta\tau_{AO} = \frac{1}{c} \cdot AO \cdot \sqrt{1-(\cos\varepsilon\cdot\cos\alpha)^2} \tag{5.4}$$

with α being the azimuth of the radio source, while for those in the east-west direction it is

$$\Delta\tau_{AO} = \frac{1}{c} \cdot AO \cdot \sqrt{1-(\cos\varepsilon\cdot\sin\alpha)^2}. \tag{5.5}$$

The accurate knowledge of the axis offset is paramount for reliable terrestrial reference frame work. It should be determined with on-site surveying methods and rotating targets. The measurements should be done at least once but preferably more often with independent approaches and distributed over the years to establish redundancy and to monitor its reliability. The reason for the necessity of such laborious and costly endeavors is that an error in the axis offset translates directly into an error in the topocentric height component. A study by Nilsson et al. (2017) has shown, that an error in the axis offset of 1 mm causes an error in station height of about 1.3 mm as a direct correlation, i.e., a positive error causes an increase in station height.

5.1.2. Construction requirements

It is needless to state that radio telescopes used for geodetic and astrometric VLBI need to be stable enough in all their components for results of highest accuracy. This applies not only for the reference point itself but also for the reflecting optics. The stability of the reference point of VLBI telescopes has been studied only occasionally (Lösler et al., 2013; Lösler et al., 2016). The results indicate that the telescope studied is stable close to the detection limit of sub-millimeter accuracy. It is worth to note that all these studies are made for *turning-head* telescopes where fixed parts of the construction can be used to mount geodetic targets. For *wheel and track* telescopes, any effect of instability, manifested as tumbling of the reference point, cannot be separated from the determination of the axis offset (Holst et al., 2019). However, if a displacement of the reference point is actually a consequence of a tilting of the telescope, this can be detected by identifying trends in its pointing model. Furthermore, a tilting process may be interpreted as a spurious telescope motion in the global frame which may be identified by comparison with velocity vectors from observations of nearby permanent GNSS installations as demonstrated on the PIETOWN radio telescope (Petrov et al., 2009).

In the construction process of a radio telescope, special care is taken that the weight of the primary reflector, the quadrupod holding the sub-reflector, and the sub-reflector are balanced by some counter-weight. The reason is that the motors and gears should be relinquished from any torques caused by imbalances for protection. For this reason, there is almost no shift in load enforced on the elevation axis, and thus on the VLBI reference point, when the telescope is tilted to different elevation angles.

The main reflector is the part of the instrument which has a great impact on the efficiency of the radio frequency reception. Often it is called paraboloid but this is only a generalization because main reflectors may also have shaped surfaces. The paraboloid hints at the fact that the reflector is indeed a rotational paraboloid which reflects the radiation, coming from a single direction, into the prime focus. Angle of incidence equals angle of emergence. In this respect, radio telescopes also obey the laws of geometric optics. The quality of the surface thus determines the gain of the reflector because divergent rays do not contribute to the focusing. As a rule of thump, at each element of the reflector, the tolerance to be adhered to depends on the shortest wavelength and is

$$\frac{\lambda}{20} = \text{tolerance} \tag{5.6}$$

For X band the tolerance thus is 1.8 mm for 8.4 GHz (X band) reception and 1.1 mm for 14 GHz, the highest projected VGOS frequency. This apparently rather crude requirement should not be mistaken for the standard deviation of surface accuracy measurements. Since standard deviation (RMS) is about one fifth of (engineering) tolerance, the requirement for this is actually 0.36 and 0.22 mm, respectively.

5.1.3. Thermal expansion of radio telescopes

Since radio telescopes are constructed with concrete and metal elements, we can expect that thermal expansion of the mechanical components leads to effects which need to be modeled in the Level-2 data analysis. The main effect is the displacement of the VLBI reference point by thermal expansion of the telescope support structure. This is mostly modeled by applying the ambient temperature to the telescope dimensions (Nothnagel, 2009) but examples exist where thermal expansion is actually measured on the telescopes directly with the help of invar rods (Johansson et al., 1996; Zernecke, 1999).

In most cases, however, the effects of thermal expansion are being modeled more rigorously by applying the expansion coefficients to the dimensions of all mechanical components of the telescopes. If we first consider only the support structure up to the elevation axis of an azimuth-elevation telescope, this normally consists of some concrete foundation and a tower produced of steel. Depending on the material, these have two slightly different expansion coefficients. Examples are γ_f of 1.0×10^{-5} $[1/°C]$ for a concrete base and $\gamma_p = 1.2 \times 10^{-5}$ for the steel tower.

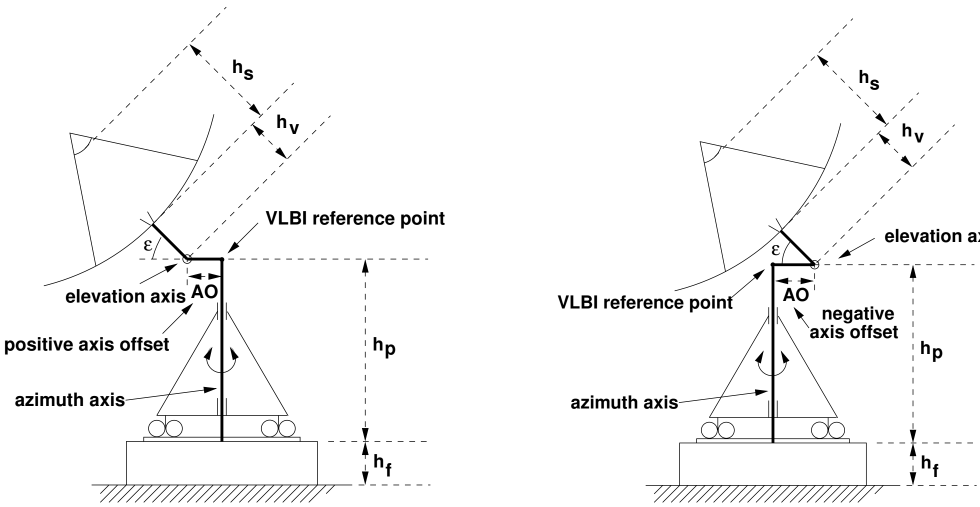


Figure 5.8: Alt-azimuth telescope mount with positive axis offset (left) and negative axis offset (right).

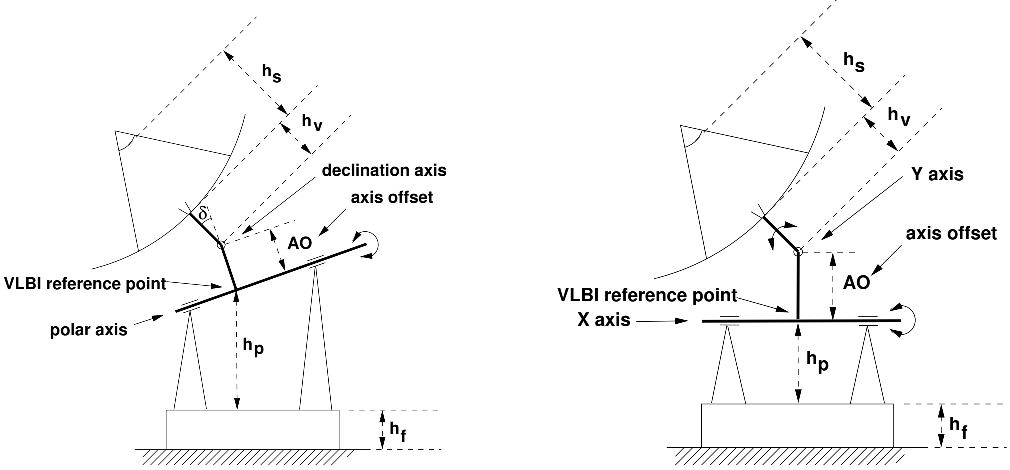


Figure 5.9: Polar telescope mount (left) and XY mount (right).

Another part of the telescope, where stability counts, is the superstructure of the telescope being responsible for the path length of the signal. In an ideal situation the total path length through the reflecting optics is considered to be stable for any direction on the sky. However, thermal expansion cause path length variations due to changes of the focal length of the telescope (Artz et al., 2014). In addition, the expansion of the legs of the quadrupod holding the sub-reflector or the feed horn in primary focus produces an extra path length at higher temperatures. Figs. 5.8 and 5.9 display and denote the components which are affected by thermal expansion. Applying a conventional reference temperature of T_0 and a time lag of $t - \Delta t$ (Nothnagel, 2009), the elevation dependent effect on the delay of telescope i can be computed according to

$$\begin{aligned} \Delta\tau_{therm.i} &= \frac{1}{c} \cdot \Big[\gamma_f \cdot (T(t-\Delta t_f) - T_0) \cdot (h_f \cdot \sin\varepsilon) \\ &\quad + \gamma_a \cdot (T(t-\Delta t_a) - T_0) \cdot (h_p \cdot \sin\varepsilon) \Big]. \end{aligned} \tag{5.7}$$

Variations of the path length due to changes in the height of sub-reflector h_f have adverse effects only within the time frame of a single session of 24 h duration because the mean effect is compensated for by the clock offset parameter in the estimation process (Sec. 13). With higher resolutions of the clock parameters down to 1 h and below, the effect is mitigated even more and the level of remaining path length variations is generally less than 1 mm. They can be modeled if all dimensions and material properties are known (Nothnagel, 2009). Consequently, variations in the height of the VLBI reference point due to thermal expansion have the biggest effect. For a telescope of 20 m diameter, the height of the elevation axis is roughly 12 m causing a height variation effect of almost 3 mm originating from a 20° C temperature difference between winter and summer.

5.1.4. Gravitational telescope deformations

The superstructure of the telescope is also responsible for path length variations of the signal due to gravitational deformations. Here, also changes of the focal length of the telescope and of the distance between the vertex of the main reflector and the sub-reflector or prime focus receiver due to changes in the gravitational load produce extra path lengths. Besides this pure geometric effect, the focal characteristics may be disturbed slightly leading to phase noise from different sections of the reflecting surfaces but these are estimated to appear only in the SNR budget and have to be neglected.

The gravitational deformations of the paraboloid and the sub-reflector create different path lengths in general but also depending on the radial distance from the optical axis. They are more critical than those of thermal origin because they are purely elevation dependent and thus change from observation to observation. As has already been investigated by (Clark and Thomsen, 1988), for a prime focus telescope, the deformations can be separated in three different components which are all elevation angle ε dependent: a) the change of focal length ΔF, b) the movement of the vertex of the paraboloid ΔV, and c) the shift of the feed horn or the sub-reflector in radial direction ΔR (Fig. 5.10). The extra path length ΔL_{grav} can then be composed of

$$\Delta L(\varepsilon) = \alpha_F \Delta F(\varepsilon) + \alpha_V \Delta V(\varepsilon) + \gamma\, \alpha_R \Delta R(\varepsilon). \tag{5.8}$$

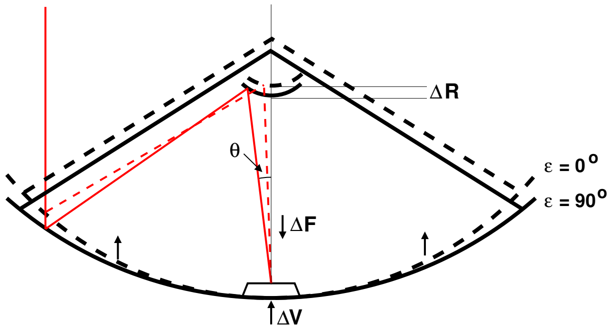


Figure 5.10: Gravitational deformations of telescope optics: Solid lines represent geometry at 90° elevation angle, dashed lines that at 0°. ΔV is the shift of the vertex of the telescope, ΔF is the change in focal length, ΔR is a shift of the sub-reflector (in case of a secondary focus telescope) or of the feed horn (in case of a primary focus telescope).

The coefficient γ is equal to 1 for prime focus telescopes and 2 for secondary focus telescopes. The net effect of the elevation-dependent displacements $\Delta F, V, R$ on the extra path length depends on scaling coefficients $\alpha_{F/V/R}$. Here, the scaling coefficients α_F and α_V according to Clark and Thomsen (1988) are linearly dependent on α_R, i.e.,

$$\alpha'_V = -1 - \alpha'_R \tag{5.9}$$

$$\alpha'_F = 1 - \alpha'_R. \tag{5.10}$$

where the V', R', F' indicate that these are parameters for prime focus telescopes. For Cassegrain systems as a type of secondary focus telescopes, (Abbondanza and Sarti, 2010) developed the relationships

$$\alpha''_V = -1 - 2\alpha''_R \tag{5.11}$$

$$\alpha''_F = 2 - 2\alpha''_R. \tag{5.12}$$

The initial concept for the computation of α_R by Clark and Thomsen (1988) assumed that all of the reflecting area of the main reflector contributes to the total change of path length in the same way. However, feed horns are always constructed in a way that they have only a limited aperture angle matching the dimensions of the reflector system to avoid unwanted spill-over from beyond the edge. Since this cannot be realized purely binary, i.e., homogeneous for the area of the reflector system and zero beyond, the feed horns are constructed with an edge taper. This realizes a roughly exponential drop of sensitivity towards the aperture angle needed to illuminate the reflector system (Fig. 5.11)

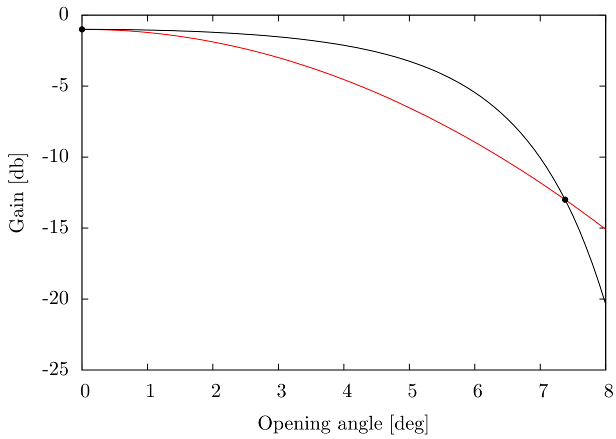


Figure 5.11: Edge taper characteristics, a) as an exponential (black) and b) as a $\sin^2$ function.

The function, which describes the drop in sensitivity depending on the aperture angle θ (Fig. 5.10), is called the illumination function and in essence serves as a weighting function for the aperture of the telescope depending on the radial distance from optical axis, i.e., signal paths near the optical axis have a much higher weight than those near the rim of the main reflector.

The determination of the exact illumination function heavily depends on measurements of the gain at multiple aperture angles of the feed horn which can be easily transformed into metric separations of the point of incidence on the main reflector from the optical axis. Unfortunately, for most telescopes only the relation between the gain on-axis and at the edge is available. With only two points of the function, this leaves room for interpretation in the sense that the shape of the function is at the discretion of the analyst with Gaussian, binomial and cosine-squared models at hand (Abbondanza and Sarti, 2010; Artz et al., 2014). Since this is a fairly new area of research, any further studies on this need to request more detailed measurements of the illumination functions on site.

The coefficient α'_R for a prime focus telescope, which scales the shift of the feed horn in primary focus (Clark and Thomsen, 1988), is

$$\alpha'_R = 8\pi f^2 \int_{t_1}^{t_2} I_n(t) \frac{1-t^2}{1+t^2}\, t\, dt \tag{5.13}$$

where $t_1 = r_1/2F$ and $t_2 = r_2/2F$ with the radial distances from the vertex r_1 and r_2, measured in the aperture plane. The distance r_1 generally is non-zero because most radio telescopes have some sort of feed horn housing at the vertex which needs to be excluded. I_n is the normalized illumination function.

For a secondary focus Cassegrain system, Abbondanza and Sarti (2010) derived

$$\alpha_R'' = 2\pi(F^2 - a^2)^2 \int_{t_1}^{t_2} I_n(t)\, \frac{t}{t^2F^2 - a^2}\, t\, dt \tag{5.14}$$

with $t = r/2F$ for the scaling coefficient, again applying a weighting through the normalized illumination function I_n. The parameter a is the reference semi-major axis of the ellipsoid at 90° elevation.

For a Gregorian system, Artz et al. (2014) adopted a similar derivation but with an implicit illumination function on the basis of a cosine-squared model. Here the sub-reflector is a triaxial ellipsoid with its longest extension in the z direction and with the first two semi-major axes (a and b) being identical. If the sub-reflector is displaced, the travel time increases by a factor of

$$\alpha_R'' = \int_{r_1}^{r_2} k \cdot 10^{\frac{c_0 + c_1 \cdot \cos^2\theta}{10}} \cdot \frac{1}{2}(D_2' + D_3 - 2a)\, dr \tag{5.15}$$

with the normalization coefficient k (Abbondanza and Sarti, 2010). D_2' the distance between the focus of the sub-reflector and point of incidence on the sub-reflector and D_3 the distance between the point of incidence and the focal plane of the secondary focus (Fig. 5.12). The aperture angle θ can easily be deduced from r. c_0 and c_1 are the coefficients of the $\cos^2$ function.

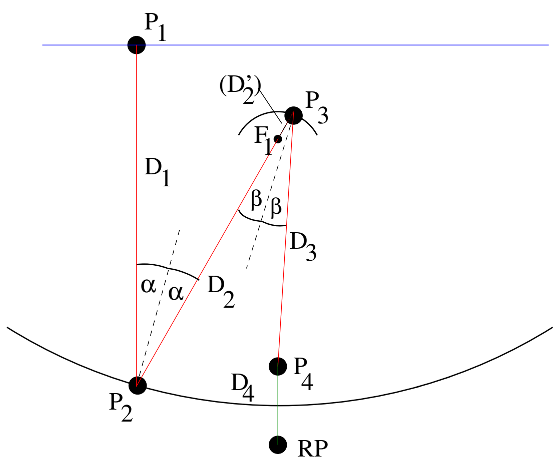


Figure 5.12: Components of path length in telescope optics.

Every telescope has its own set of coefficients $\alpha_{F/V/R}$. A rough number for $\alpha R''$ for a secondary focus telescope is 0.9 (1.06 for Effelsberg (Artz et al., 2014) and 0.8 for Medicina or Noto (Abbondanza and Sarti, 2010)) which, according to Eq. 5.11 and Eq. 5.12, lead to $\alpha V'' = -2.8$ and $\alpha F'' = 0.2$. This indicates that changes in the position of the vertex of the paraboloid relative to the VLBI reference point have about a ten times larger effect on the signal path length than changes in the focal length and 3 times larger an effect than of changes of the sub-reflector position.

Similar relations also apply for primary focus telescopes where Abbondanza and Sarti (2010) quote an $\alpha R'$ of about 0.7 resulting in an $\alpha V' = -1.7$ and an $\alpha R' = 0.3$ according to Eq. 5.9 and 5.10. Here, the multiplication factor of a shift of the vertex of the paraboloid is about a factor of 2.5 larger than that of the feed horn $\alpha R'$.

Although variations of ΔR may originate from several phenomena, we only have to consider line-of-sight shifts of the sub-reflector or feed horn (Fig. 5.10) in the telescope-fixed reference frame. Position variations predominantly happen due to gravitation when the telescope is tilted, e.g., by bending of the struts holding the receiver (Sarti et al., 2009). These need to be measured locally. Presently, the most suitable technique for measuring the deformations of the primary reflector resulting in a ΔF is terrestrial laser scanning (TLS). Performing these at discrete elevation angles between zenith and horizon provide reliable estimates of the changes in focal length (Artz et al., 2014; Holst et al., 2015; Sarti et al., 2009). In some cases, the other parameters (ΔR, ΔV) can be deduced from the TLS results (Artz et al., 2014) but often they need other sensors mounted on the structure of the telescope (Bergstrand, pers. comm.).

5.2. Receiving systems

Today, geodetic VLBI receiving systems come along in two different versions, legacy systems for simultaneous S and X band observations and VGOS broadband systems. Since the technical developments of the VGOS system are making use of modern electronics components, there has been tremendous progress in increasing the observed bandwidths. However, the general background of radio telescopes and the receiving equipment have remained the same. As can be seen in Fig. 1.2, the basic elements of the first stage of the receiving system called front-end consists of radio frequency amplifiers, local oscillators and mixers. These elements are mostly mounted directly in the telescopes themselves. The back-end consists of samplers and recording units, which are normally located in the control room/building. The information related to the front- and back-ends are separated in three sub-sections, general receiving concepts (Sec. 5.2.1), legacy system hardware (Sec. 5.2.2), and VGOS system hardware (Sec. 6.3).

5.2.1. General receiving concepts

The actual antenna element of a radio telescope is a feed horn, basically a funnel in which the electro-magnetic radiation is collected and coupled into electronics channels. From the injection into the feed horn, the quasar signal is lead into the first amplifier stage consisting of low-noise amplifiers (LNA), e.g., field-effect transistors (FET). These are enclosed in a cryogenic dewar (Fig. 5.23) which means that the air is replaced by helium cooled down to about 40° K.

Cooling is necessary to reduce the thermal noise of the electronic amplifier components to a minimum to avoid spoiling the quasar noise by thermal noise. In the case of the VGOS broadband feeds, even these themselves are cooled. Although there will be further system noise sources down the receiving and digitization line, the noise reduction of the first amplifier stage determines the final receiver noise temperature which represents the total noise introduced by the receiving system in units of equivalent temperature. Knowing all noise power contributions of the receiver components, T_{rec} could be computed by

$$\frac{P_{\mathrm{N}}}{B} = k\, T_{rec} \tag{5.16}$$

where P_N is the noise power in Watt, B is the total bandwidth in Hz over which the noise power is measured, k is Boltzmann's constant (1.38×10^{-23}) Ws/K producing the receiver noise temperature T_{rec} in degrees Kelvin. Unfortunately, this is not the only contribution to the total system noise. Also the noise of the ground T_{grd} to be limited by the edge taper of the telescope and of the sky T_{sky} which is well known as the cosmic background radiation of the sky of 2.726 K. Consequently, the system temperature sums up to

$$T_{sys} = T_{rec} + T_{grd} + T_{sky} \tag{5.17}$$

Although one could compute the receiver temperature by Eq. 5.16, in practice the system noise temperature is determined by switching on and off a noise diode with a calibrated noise temperature output T_{cal} placed in front of the first amplifier stage and computing the relationship

$$T_{sys} = \frac{T_{cal} \cdot P_{\text{cal-off}}}{P_{\text{cal-on}} - P_{\text{cal-off}}} \tag{5.18}$$

read from a total power integrator as $P_{\text{cal-off}}$ and $P_{\text{cal-on}}$. For this, the telescope has to point to a part of the sky where the overall power is at a minimum.

In the absence of a calibration of the noise source, one can also determine T_{cal} by pointing on and off a well known calibrator source. Here, the same concept of Eq. 5.18 applies as it does for pointing checks where a 5-point sequence is followed (on-source, first null above, first null below, first null east, and first null west). The system temperature is needed for precision considerations (Sec. 3) and for determining necessary scan lengths in the scheduling process (Sec. 9).

The output of the first stage amplifiers, the radio frequency radiation in the GHz regime has been too high for direct digitization. For this reason there is a need for down-conversion. Mixing down the RF frequency to some MHz can be considered as subtracting a carrier wave as commonly used in terrestrial electronic distance meters. The down conversion is done in multiple stages and at this stage frequency stability comes into play. Although these stages may be realized differently at the various telescopes, the entire down conversion should be phase coherent in the receiving chain of a telescope but should also be coherent to other telescopes. It should be mentioned that there are present day developments of direct sampling of GHz frequencies which may cope with these frequencies, but these are still in an experimental stage.

The driving element of the frequency down conversion is the telescope's frequency standard which almost always is a hydrogen maser (H-maser, Fig. 5.13) because it guarantees a frequency stability of 5×10^{-14} in 50 minutes (Nothnagel et al., 2018). A hydrogen maser normally has a 5 MHz and a 1 pps output (1 pulse per second). The 1 pps signal is used to drive the station clock which is embedded in the formatter or sampler electronics. This produces the time tags in the data framing. The 1 pps signal needs to be synchronised to UTC as best as possible but it also helps to know and document the deviations from UTC. Often the deviations are determined with GPS time transfer equipment and stored in the telescope's log file as "fmout - gps" meaning formatter output reading minus GPS time tag. This information is helpful to the correlator centers to speed up operational correlations by eliminating the need for a dedicated fringe search. For this to work, the maser offset with respect to UTC needs to be accurate only at the 100 ns level. A de-facto synchronization of the clocks of the radio telescopes takes place in the level-2 data analysis where clock offsets are estimated with picosecond precision (For definitions of the data analysis levels see Sec. 15.3).

Figure 5.13: Hydrogen maser frequency standard.

The 5 MHz output of the H-maser (may also be 10 MHz for newer versions) is the driving signal for all local oscillators used for frequency conversion, phase calibration and data framing in the sampler. Of the same importance as the stability of the frequency standard itself is the quality of the distribution of the frequency to the individual pieces of equipment. Here, phase coherence has to be guaranteed to keep up with the high demand for coherence of the the individual down conversion steps at a telescope but also between the signal streams of the two telescopes forming a baseline.

The main concept of down-conversion is that the entire frequency band, which is received, is first separated into several sub-bands (8 at X band, 6 at S band, and 32 in the final VGOS mode). These sub-bands are then down converted to baseband for each sub-channel individually, which means that the bandwidth of each of the sub-channels now ranges from 0 to F MHz, therefore the name baseband. F was 2 MHz at times of the early Mark 3 system, increased to 4, 8, or 16 MHz in later Mark 4/5 developments, and is 32 MHZ for the current VGOS system (Niell et al., 2018). Since the signals of each sub-band run through separate electronic circuitry, each processing chain can also be denoted a data flow channel, or briefly channel as it was common use in the legacy system era. So, beware that there may be some inconsistencies in the usage elsewhere.

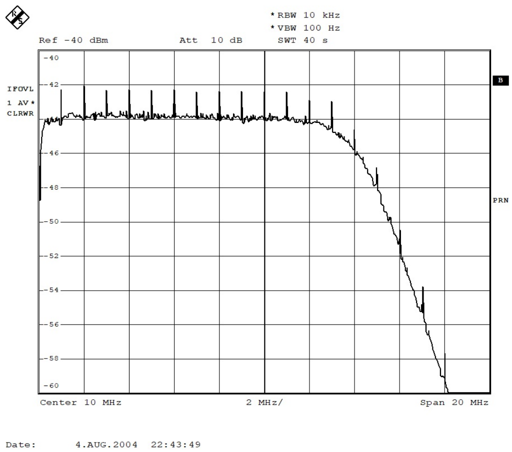


Figure 5.14: Bandpass shape. Courtesy Gerhard Kronschnabl.

When looking at legacy analog systems, a critical element was the shape of the bandpass, i.e., the form of the power pattern with respect to frequency. This depends on the quality of the filters and their phase stability. The sharper the filter (for amplitude), the less stable are the frequency dependent phases. This is due to the Kramer-Kronig-relationship which relates real and imaginary component of meromorphic functions. Fig. 5.14 depicts a real 16 MHz bandpass of a Mark IV baseband converter with a drop of amplitude of -8 dB at 16 MHz and still -16 dB only at 20 MHz. Beyond 16 MHz the power should be zero but this was never the case and this excess power also got folded into the digital domain.

Fortunately, today with digital baseband converters such as DBBC or RDBE, the problem is minimized. Digital filters have better characteristic but also not 100%, though.

Today, a serious problem is radio frequency interference (RFI) through strong artificial emitters which affect the receivers especially those of S band. In particular mobile phone transmitters but also WiFi produce so much radiation that in some cases one or more of the sub-bands may saturate the receiver and turns them unusable. This is the reason, why VGOS frequency bands start at 3.0 GHz. In legacy systems, heavy RFI is tried to be avoided by sub-band re-allocation but the multitude of transmitters at the different telescopes drives this endeavour futile. For this reason, digital filtering (notch filters) are developed to be used in the fringe fitting process. However, RFI is always a nuisance and reduced the SNR.

The next element in the receiving chain is the digitization of the individual sub-bands which is done according to the Nyquist theorem with two samples of one bit each per Hz (Fig. 5.15). For any sampling to work, a sampling signal of twice the frequency to be sampled has to be employed generating a sampled analog signal (Takahashi et al., 2000). By pure translation of the voltages into individual bits, digitization is realized. "0" represents a negative voltage and "1" a voltage of zero or larger. This is called 1-bit sampling although 2 bit (or 2 samples) are needed for one cycle. In other words, the first sample averages the voltage of the 0 - 180° part of the cycle and the second sample the 180 - 360° part.

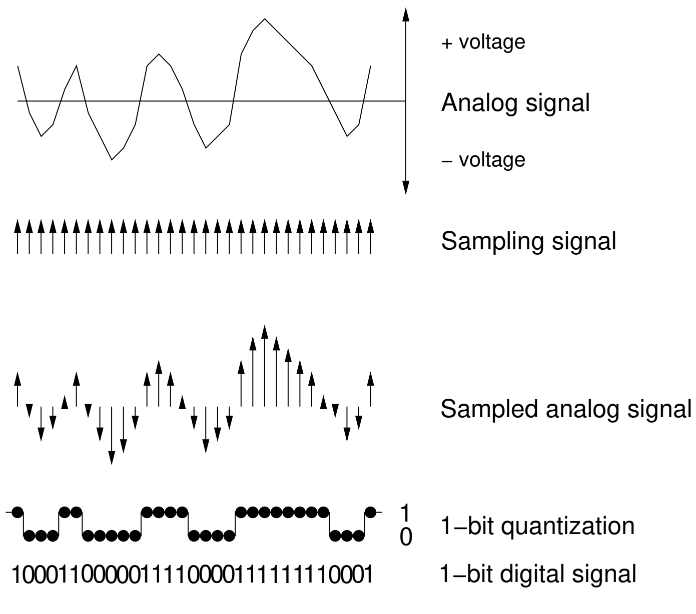


Figure 5.15: 1 bit digitization (adapted from Takahashi et al., 2000).

In many setups of observing sessions 2-bit sampling is applied. Here, a second bit is added to each sample indicating whether the voltage is below or above a (half-power) voltage threshold (amplitude bit, Fig. 5.16). For amplitude calibration of the sampler, modern digital baseband converters such as the DBBC (Digital BaseBand Converter) or the RDBE

(Reconfigurable Open Architecture Computing Hardware [ROACH] digital backend) do a constant statistical test on the amplitude bit querying whether the nominal distribution of 16% above the positive amplitude threshold or 16% below the negative amplitude threshold is met. Depending on the outcome, the voltage thresholds are shifted up or down to keep the expected bit distribution intact. This may be done with repetition rates of 1 Hz or even below. It should be mentioned here that the amplitude information can be sampled with finer resolution requiring 8- or even 16-bit representation. 8-bit sampling is employed in the RDBE and commonly used in VLBI space navigation systems.

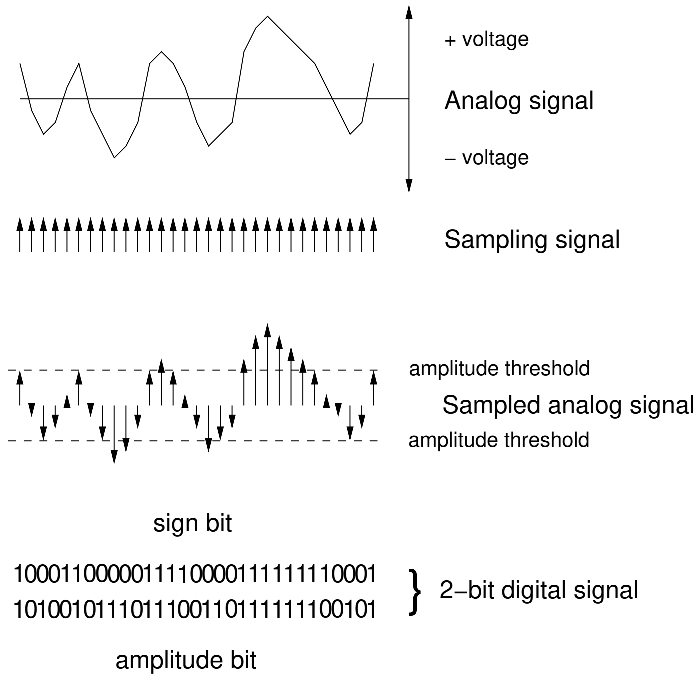


Figure 5.16: 2 bit digitization.

In some references, fringe plots (Appendix F.1), or control files such as the .vex files for telescope and correlator control, the sample rate is indicated. For example, for 8 MHz wide channels/sub-bands, the sample rate normally is 16 Ms/s (Mega samples per second) which is derived from the two bit sampling interval of a one Hz cycle. No information is given by the sample rate whether the observations are carried out in 1-bit or 2-bit mode. This is listed in the $TRACKS section of the .vex file as separate storage "tracks" for the sign and magnitude bit. In *fourfit* fringe plots, the information is given as "bits/sample".

Digitization and storage are mostly done in a real data mode with single sideband. Another way of treating the time varying data is in complex mode . One way of doing that is a modulation at the front end. Here, the analog signal is superimposed twice by the same local oscillator frequency, one in-phase and one phase-shifted by 90°, producing I (in-phase) and Q (quadrature) components (Fig. 5.17). This mechanism produces a double sideband (DSB), i.e., LSB and USB, representation with local oscillator frequency defining the reference frequency, which is now in the middle of the band and not at the edge for single sideband (SSB) data. This has consequences for correlation because it needs an extra phase shift (Chris Phillips, priv. comm.).

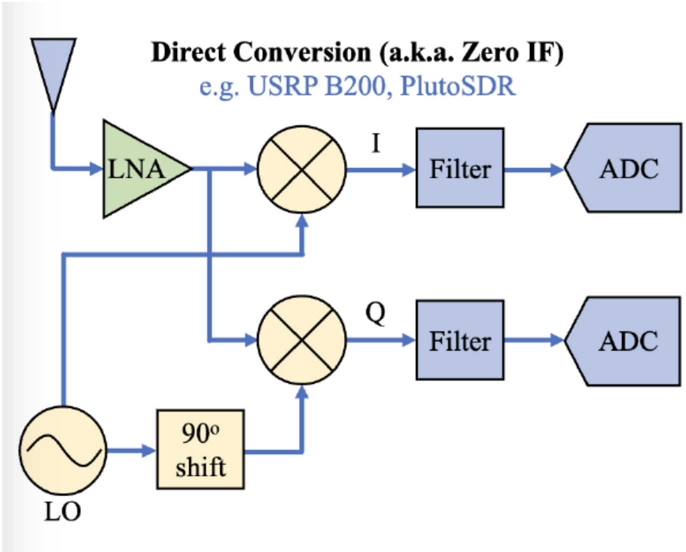


Figure 5.17: Generation of complex data by phase-shifting the local oscillator (LO) frequency and recording the in-phase and quadrature (I/Q) components (Figure courtesy Chris Phillips).

In terms of storage, today, the VDIF format (VLBI Data Interchange Format) is commonly used for recording and transfer of VLBI raw data[3]. It was proposed in 2009 as a result of an internationally constituted VDIF Task Force appointed in June 2008 to study problems of incompatible data and create a recommended uniform transport-independent VLBI data-format standard. Before that time, a number of different formats existed which had been devised for specific VLBI recording hardware. The recording tape drive families of Mark III, Mark IV and K3 as well as the disk recorders Mark 5/6, and K5 all had their own formats with the consequence that the correlators needed matching playback units for a correlation of these data. Since especially the tape recorders and playback units were rather expensive, it took long before data with incompatible formats could be correlated. This is overcome now by VDIF and recording on magnetic disk drive RAID units such as Flexbuff.

Just for completeness, I should mention two other formats, the Raw Data Exchange Format (RDEF) and the DSN DVP format. The first one was devised to have a recommended standard format for use in exchanging Delta-DOR raw data among space agencies. The second was an interim format of the DSN antennas before they also switched to VDIF (Simone Bernhart, priv. comm.).

In the context of digitization, there sometimes is mention of a Nyquist zone with the attribute "first, second, ...". This has to do with the fact that one always needs sampling

[3]https://vlbi.org/vlbi-standards/vdif/

with a certain clock rate which has to be as minimum double the frequency bandwidth to be digitized. If this band is for example 100 MHz wide but the clock rate is only 50 MHz (20 ns), actually only a bandwidth of 25 MHz could be sampled providing an unambiguous representation. This could be the slot of 0 to 25 MHz which is the First Nyquist zone. The slice of 25 to 50 MHz then is the Second Nyquist zone of this frequency/sampling rate relation and so forth. This Second Nyquist zone could be sampled with the same 50 MHz (20 ns) clock rate but only providing to submit to the sampler input the band 25-50 MHz, or still 50-75 MHz for the third Nyquist zone, or finally 75-100 MHz for the fourth Nyquist zone in the 100 MHz wide band to be digitized. Allocation of the 25 MHz wide sub-bands into separate input bands, e.g., into the DBBC, needs the respective filters to allow the necessary distinction of the sub-bands in the hardware (G. Tuccari, priv. comm.).

5.2.2. Legacy S/X band system hardware

The description of the legacy system hardware should be considered as generic because there are a number of different realizations (Fig. 5.25). We start again with the front-end. Legacy systems mostly use concentric feed horns for two bands in the radio frequency (RF) domain, X band as primary observing frequency and S band for calibration of ionospheric refraction. Concentricity of the feed horn for many frequencies is important for pointing the telescope for simultaneous reception of all frequencies to be observed and processed. The focal points of the different frequencies have to lie on the same optical axis. Within limits, it does not matter whether they are slightly displaced on the common optical axis. Any resulting path length differences are absorbed in the clock parameter estimation. However, the reflection of the radiation from the main reflector or the sub-reflector should collimate in a single focal point. If the feed horn is not constructed properly, any disagreement of the focal points will result in a deterioration of the antenna gain.

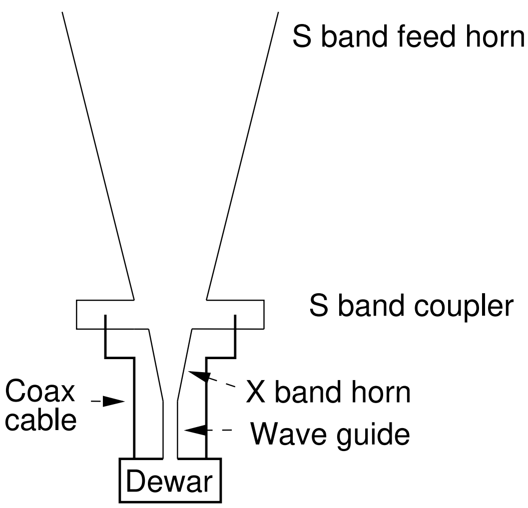


Figure 5.18: Schematic concentric feed horn for S and X band reception

In general terms, the entire system consisting of first stage amplification and heterodyning to lower frequencies is called the receiver. Although other realizations certainly exist, a concentric feed horn for simultaneous S and X band reception may resemble a big funnel as depicted in Fig. 5.18 for a secondary focus antenna. The funnel collects all frequencies but is optimized for S and X band. The top part (Fig. 5.19) is designed for optimal S band and the bottom part (Fig. 5.21) for X band reception (In fact, the entire S band horn also consists of a second cone element with narrower aperture below that on the photo). The S band signal is decoupled from the waveguide element into coaxial cables with four S band couplers (Fig. 5.20) matching the 13 cm wavelength and mounted cross-wise around the bottom part of the S band horn. The little pin visible in Fig. 5.20 is the actual antenna element. The reason for having four of them is to produce right hand circular polarization (RCP) through phase-shifting the four inputs by $\pi/2$ each. From there, the signals are led into the amplifier chain by coaxial cables. The X band radiation remains unaffected in the S band part of the horn and is fed into the receiver through wave guide elements at the bottom of the feed directly (Fig. 5.21), again matching the 3.4 cm wavelength. A $\lambda/4$ plate in the path produces RCP.

Figure 5.19: Top part of S band horn.

Figure 5.20: S band waveguide to coaxial coupler.

Figure 5.21: X band feed horn

Both horn components are machined with corrugations. The purpose of these is to produce the characteristic beam of the horn which should illuminate the reflector (in this case the sub-reflector) with maximum gain everywhere but none beyond the rim. This is not strictly possible but is approximated with the edge taper to minimize the gain as much as possible towards the rim (Fig. 5.11). Feed horns are specifically constructed for a radio telescope where the f/D relationship (focal length over diameter) determines the width (or narrowness) of the feed cone. For large f/D and consequently for secondary focus telescopes, the feed is narrower than for prime focus telescopes.

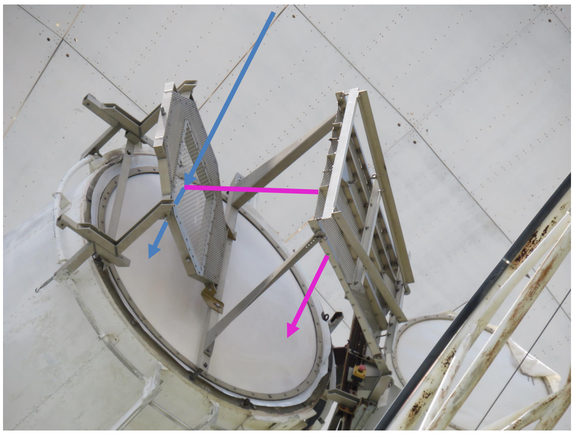


Figure 5.22: HartRAO dichroic mirror assembly on top of cone containing feed horns of several frequencies (left). The reflected S band signal is depicted in magenta (right). S and X band horns are located underneath the mylar lid of the cone. Photos by Marisa Nickola.

It should also be mentioned that some receiver installations use dichroic mirrors to reflect part of the radiation (often the longer wavelengths) in the direction of a separate feed horn (for S band) before it reaches the horn for X band. The additional S band reflections lead to a slightly reduced loss in gain, which is acceptable, because the S band observables have reduced quality requirements compared to that of the primary frequency of X band. Fig. 5.22 shows the assembly of the 26 m HartRAO telescope which is taken off for astronomical single frequency observations to avoid blockage of other feed horns in secondary focus.

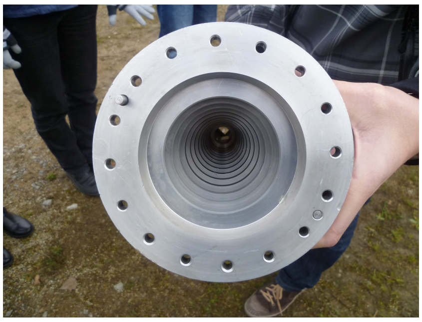


Figure 5.23: Top view of X band horn.

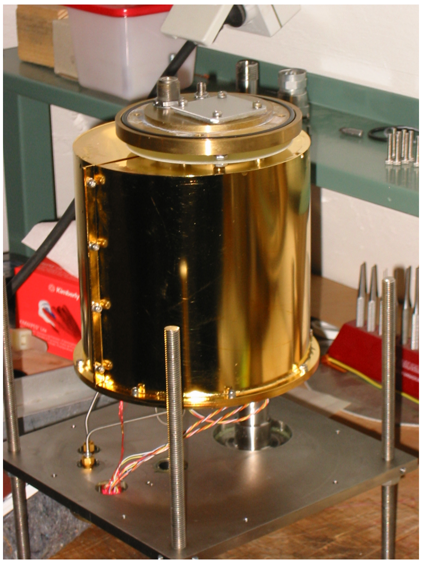


Figure 5.24: Cryogenic dewar containing low noise amplifiers.

The cryogenic dewar of legacy telescopes generally contains a first stage amplifier and a mixer for both frequency bands separately. The purpose of the mixers is to bring down the radio frequencies (GHz domain) to about 500 MHz to 1000 MHz called intermediate frequencies, IF. This is done by mixing the RF frequencies with some local oscillator frequency, e.g., 2020 MHz for the S band signal and 8080 MHz for X band (Fig. 5.25). Only the frequency difference components are of further use while the inevitable sums are filtered out.

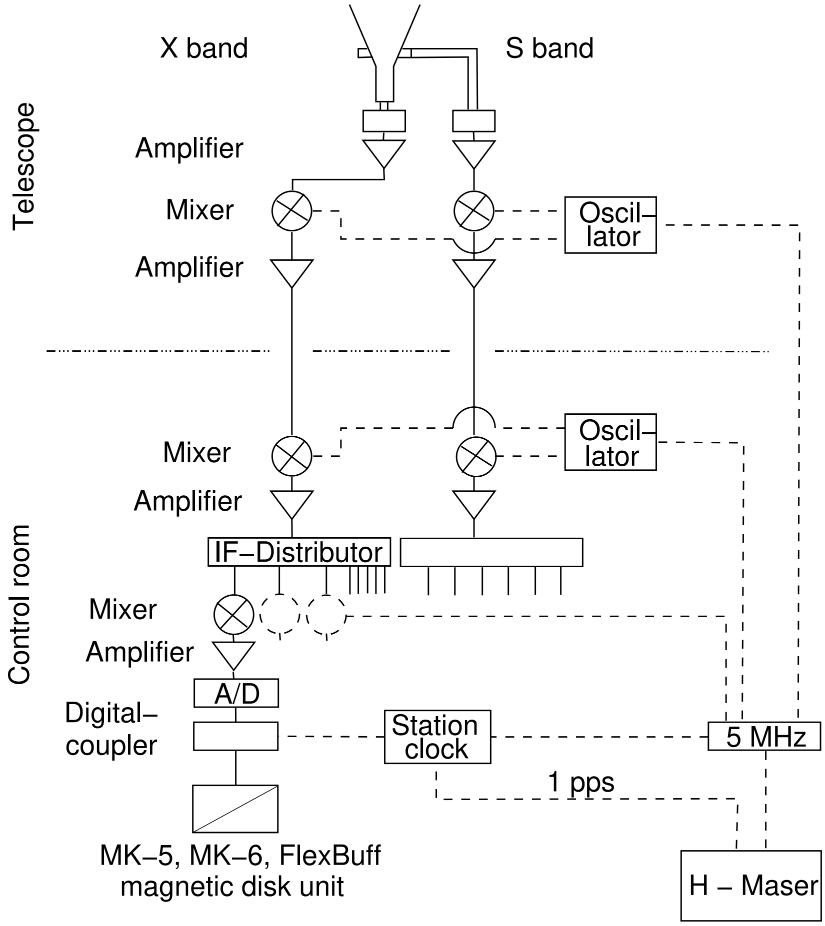


Figure 5.25: Schematics of RF to digital data flow in a VLBI system.

For further processing and registration, the IF signal is amplified again and brought down from the receiver equipment in the telescope to the control room either in coaxial cables or nowadays also in fibre optic cables. The S and X frequency bands in legacy telescopes have always been transported via separate cabling.

In the control room, an IF distributor helps to separate the frequency bands into several sub-bands and the signals are down converted to baseband for each sub-channel individually as described in Sec. 5.2.1. Along the same lines as above all sub-bands are digitized and time tagged.

In the early days, the processing chain was analog down to the analog to digital (A/D) converter which was called the formatter. Today, with the goal to operate digitally as early in the receiving chain as possible, the IF distributor, the last mixer stage, A/D converter, and sampler are included in digital baseband converters, such as the DBBC or the RDBE family.

The distribution of the sub-bands within the spectrum of a frequency band follows the ideas of Rogers (1970) for bandwidth synthesis which allows spanning as much bandwidth as possible with limited recording capabilities (Sec. 2). For the description of a frequency sub-band, in most cases, one of the band limiting frequencies is given together with the indicator upper sideband (USB) or lower sideband (LSB). Throughout the VLBI community, the sub-band reference frequencies 2225.99, 2245.99, 2265.99, 2295.99, 2345.99, 2365.99 MHz of S band and 8212.99, 8252.99, 8352.99, 8512.99, 8732.99, 8852.99, 8912.99, 8932.99 MHz of X band, all with USB are currently allocated. At X band, two more bands are often used at 8212.99 MHz (LSB) and 8932.99 MHz (LSB) to add two more sub-bands to the historically provided 14 sub-bands of the early Mark III electronics and tape recorder era. Fourteen was the number of tracks recordable on the first video tapes of the Mark III system which was overcome by the "high density tape upgrades" which allowed 16 tracks. Since there were only 14 baseband converters (BBC) in the Mark III electronic rack, the lower sideband output of the first and last sub-bands was employed to make use of the extra two tracks and increase SNR by this additionally recorded data.

The Mark 5 recording units in use today consist of a chassis, which is normally mounted in some electronics rack, and exchangeable, transportable disk units, each containing eight commodity disks for storage of the data, thus Eight Packs (Fig. 5.26). The latter are used for intermediate storage and later transmission of the data by Internet lines or for shipment to the correlator by courier service. These are known as Mark5 recorders. Alternatively, also RAID storage systems are in use (called Flexbuff[4]). Naturally, they are mainly used for transportation of the data to the correlators via the Internet.

Depending on the number of individual sub-bands and their bandwidths, the total data volume for legacy systems may be up to 1 GigaBit per second (Gbps) or 4.3 TerraByte (TB) per telescope at a 40% recording cycle (the rest is used for slewing of the telescopes).

Figure 5.26: Transportable disk pack of Mark5 system (Eight pack).

[4]https://aaltodoc.aalto.fi/bitstream/handle/123456789/16026/master_Salminen_Tomi_2015.pdf

6. VLBI Global Observing System

The VLBI Global Observing System (VGOS) is the name for the current developments and realizations of a modern geodetic VLBI observing concept. Compared to the legacy system of S/X band observations, it has a number of new features which are described in this chapter.

6.1. History

The initial project of VGOS started off in 2003 with the establishment of the IVS Working Group 3 (WG3) called "IVS Working Group on VLBI2010" (Sec. 15.3). The purpose of this working group was to "produce a report with recommendations for a new generation of systems that meet the following criteria: (a) Highest-precision geodetic and astrometric results, (b) Low cost of construction, (c) Low cost of operation, and (d) Fast turnaround of final results."[5]. What followed was a time of extensive systematic simulations and intense discussions on all effects and all technical compartments affecting the quality of geodetic and astrometric VLBI results.

In September 2005, the IVS Directing Board approved the final report of WG3 named "A Vision for Geodetic VLBI – VLBI2010: Current and Future Requirements for Geodetic VLBI Systems." (Niell et al., 2006). The most important findings in the context that follows were reflected in the part that was called the "3.1 Design a New Observing System Based on Small Antennas". In the following years, the new concept led the way of the IVS into the future and triggered many detailed investigations such as the design concept of the VLBI2010 system (Petrachenko et al., 2009). With the decision of the IVS Directing Board at the Royal Observatory of Madrid, Spain, on March 9, 2012, the name of the project was changed from VLBI2010 to VLBI Global Observing System (VGOS) to take into account that the year 2010 had passed and that a permanent name should have a relationship to the Global Geodetic Observing System (GGOS) of the International Association of Geodesy (IAG) where 1 mm coordinate and 0.1 mm/y velocity accuracy targets had been demanded. Since then, the establishment of the VGOS network and related processing capabilities have been on the agenda of many IVS components. A fair number of new telescopes have been built and reached operational status while correlator capacity and post-processing software have been developed. Between December 2014 and January 2017, the first VGOS observing sessions weren observed and successfully analyzed proving the validity of the concept (Niell et al., 2018). With VGOS-imminent new technologies needing extensive research, developments of VGOS will provide important topics for many years to come.

[5]https://ivscc.gsfc.nasa.gov/about/wg/wg3/index.html

6.2. Technical framework

The simulations carried out in the VLBI2010 project led to the insight that the most important area hindering progress with the existing (legacy) S/X band system was, and still is, the refractive behavior of the microwaves in the atmosphere. Improvements of the geodetic and astrometric VLBI results could, thus, be achieved only if the neutral atmosphere and here the influence of water vapor, manifested in the extra wet atmosphere path length, is determined with an accuracy approaching the mm level. Since the wet atmosphere, as it is commonly called, cannot be modelled or measured with such an accuracy, estimating zenith wet delays (Sec. 13.7.2 and 13.7.3) is the current method of choice.

And here is where the "small antennas" come into play. It was beyond dispute that upgrading the many dual-use (astronomical and geodetic) radio telescopes and aging equipment of the purely geodetic ones was required. Only with less mass to accelerate, move, and decelerate can the slewing speed between radio sources be increased with mechanically tolerable strain. Consequently the time between observations of different regions of the local sky is reduced allowing for more observations per time unit. With these, the sampling of the atmosphere above each telescope can be performed much more densely than with the normally much slower legacy telescopes. Consequently, estimating the zenith wet delay (or any other parameter reflecting the state of the refractive irregularities) with much higher time resolution and better quality will be possible.

For the consequences following this insight we should recall that the signal-to-noise ratio (SNR) achieved during an observation is the positive driving parameter of the precision of the phase delay observable (cf. Eq. 3.5), and, together with the spanned bandwidth (Eq. 3.12), also that of the precision of the group delay observable (Eq. 3.11). According to Eq. 3.3 (SNR $\propto \sqrt{S A_{(1)} A_{(2)} B T}$), the SNR is dependent on the effective collecting areas of the two telescopes forming the baseline, $A_{(1)}$ and $A_{(2)}$. Reduction of one of these parameters (compared to the legacy telescopes of 20 m diameter) by a factor of 2-3, must be compensated by an increase in one or more of the other parameters.

However, the correlated flux density of the observed source S should be kept below or around 1 Jy to avoid an increase in probability of adverse effects by the structure of resolved sources. Likewise, the integration time T should be kept small for a fast sequence of observations. This excludes these two parameters as candidates for compensation for the reduced collecting areas of the agile, small diameter VGOS telescopes and consequently leaves only the bandwidth B to be increased if the same SNR is to be maintained. In addition, if the precision of the observables were to be improved, as an inherent part of the goal of reaching "highest precision of geodetic and astrometric results", the bandwidth B was the remaining candidate for increase to compensate for the smaller diameter telescopes.

Another early consideration for the realization of the desired high temporal density observations was the achievable shortness of the scan duration (on-source time). If the lower limit of source brightness is assumed to be 250 mJy, two twelve meter diameter telescopes can achieve an SNR of 20 within 10 s integration time (Niell et al., 2018). Prerequisites are that the recording rate is 16 Gbit/s and that the telescopes have aperture efficiencies of 50% and system temperatures of 50 K, which is equivalent to an SEFD (see Sec. 3) of 2500 Jy.

Finally, the simulations showed that the highest tolerable delay precision to achieve the 1 mm telescope position goal was 16 ps. This included the impact of all known sources of noise-like errors. From experience with the legacy S/X band systems it was clear that the signal chain would also need a re-design for the required delay precision target.

With all these preconditions in mind, the idea then was to develop a system with four bands of 1 GHz each spread over a range of 2 - 14 GHz sky frequency (Petrachenko et al., 2009). With this initial general definition, the so-called broadband system was born. The lower frequency limit of 2 GHz was chosen in order to maintain compatibility with existing S/X systems and allow for possible mixed-mode observations. Allocating a wider range of the sky frequency limits and not just a continuous 4 GHz was chosen to make use of the spanned bandwidth effect of Eq. 3.11.

The technical realization of the microwave system then faced another obstacle. While existing signal chains with feed horns for S and X band frequencies could employ so-called quarter wave plates to introduce a 90° phase shift in the polarization planes for extracting the left- and right circular polarization components of the incoming radio source signal (see Appendix C on polarization issues), there was no such solution for wide-band feed systems. These are currently only able to physically receive linearly polarized radiation. While hardware realizations of this problem were only in an experimental stage at that time, receiving and recording two perpendicular linearly polarized wave fronts seemed to be a good solution. This appears even more beneficial because the number of bits to be correlated would be doubled compared to a single circular polarization as realized in the legacy S/X system and lead to an increase of the SNR by $\sqrt{2}$ pushing the accuracy up even further, at least theoretically.

With the definition of VGOS to cover 2 - 14 GHz, the overall boundaries of the spectrum were set but that left the question of how to populate this spectral region with actual receiver bands. Ideal would be full coverage of the 2 - 14 GHz spectrum, but plenty of obstacles inhibit this. Fortunately, bandwidth synthesis (BWS) (Rogers, 1970) and its accuracy considerations, based on the recording of sub-bands of the frequency bands mentioned already above also work for the VGOS broadband system and helped again to save on hardware, storage, transmission, and processing costs. Here, the four bands are populated with 32 MHz wide sub-bands (Fig. 6.1). Current limitations force the use of discontinuous sub-band selections within a bandwidth of 512 MHz. However, the long-term goal is that each band of 1 GHz bandwidth is fully populated routinely with 32 MHz wide sub-bands/channels.

In the context of nomenclature, the application of bandwidth synthesis with all VGOS bands and sub-bands in a single process for the determination of the group delay and the

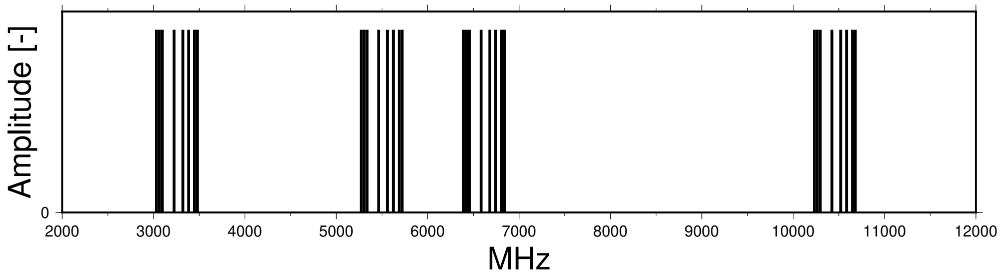


Figure 6.1: Sub-band allocation of current VGOS setup (for frequencies see Tab. 6.1), sub-band bandwidth 32 MHz.

difference in total electron content (dTEC) (see Sec. 6.4) produces a group delay which for sure is a real multi-band delay. As a consequence, care has to taken and the context needs to be observed when using the different delay expressions.

The signal chains for four bands of 1 GHz bandwidth each (called A, B, C, and D) is quite a challenge because of the wide spread of frequencies of up to 12 GHz and the technical feasibilities in the different frequency domains. At present technical limitations allow only for 8 sub-bands in each band (Niell et al., 2018). The setup as used in the current (as of 2022) IVS VGOS observing sessions is listed in Tab. 6.1 and depicted in Fig. 6.1.

Table 6.1: Upper bandedge frequencies currently in use. Each sub-band has a nominal width of 32 MHz (lower sideband).

Band	A	B	C	D
	3032.40	5272.40	6392.40	10232.40
	3064.40	5304.40	6424.40	10264.40
	3096.40	5336.40	6456.40	10296.40
	3224.40	5464.40	6584.40	10424.40
	3320.40	5560.40	6680.40	10520.40
	3384.40	5624.40	6744.40	10584.40
	3448.40	5688.40	6808.40	10648.40
	3480.40	5720.40	6840.40	10680.40

An area of research still is where to optimally locate the four bands in the broadband spectrum. The decision is still pending because the limitations of technical components, radio frequency interference, and the simultaneous estimation of the total electron content difference (see below) put severe restrictions on the optimization function. Consequently, there is a need for close communication between hardware developers, station staff, and data analysts to iterate on optimal configurations.

6.3. VGOS system hardware

The new developments for VGOS have brought up concentric feed horns which are capable of receiving, other than the S/X horns, the full range from a few GHz up to 14 GHz simultaneously. Examples are the Eleven Feed family (Yang et al., 2013) covering 1.4 to 14 GHz or the Quad Ridge Flared Horns of Cosmic Microwave Technology, Inc.[6] which nominally operate from 2.3 to 14 GHz. Eleven Feeds are installed in the Wettzell and Onsala Twin Telescopes (Fig. 6.2) while Quad Ridge Flared Horns are in use in US and Australian telescopes (Fig. 6.3).

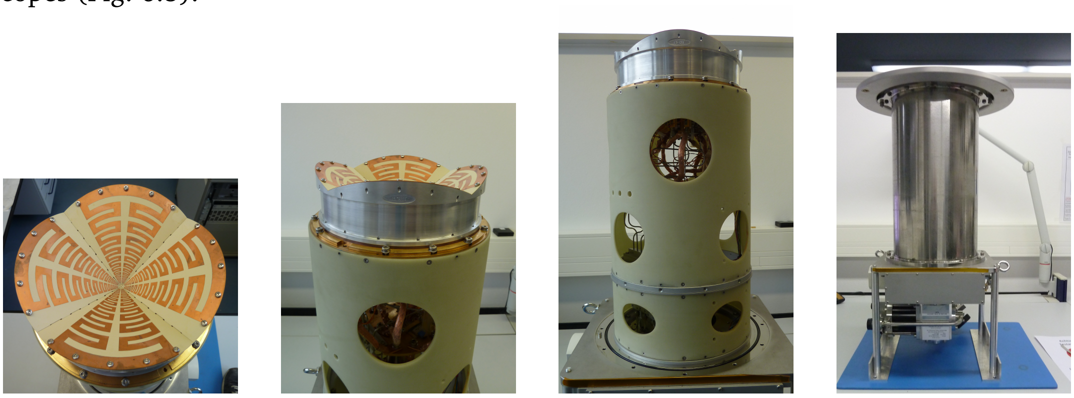


Figure 6.2: Eleven Feed for 1.6 to 14 GHz. a) as seen from top, b) with housing for amplifiers, c) with bottom plate of cryogenics, d) with outer cryogenic shielding and cryogenic pump underneath (courtesy Gerhard Kronschnabl).

Figure 6.3: Quad Ridge Flared Horn (Courtesy Cosmic Microwave Technology, Inc.).

The signal flow from the receiver down to the control room depends on the specific hardware installation. Multiple bands can be transferred with joint RF over Fibre (RFoF) systems or may be split into two or more downlink cables. For example, in the US telescopes the LNA output is separated into two chains, one for the lowest band and another one for the three higher bands together (Niell et al., 2018). This is duplicated for both polarizations (Fig. 6.4). The reason is avoidance of potential "RF over Fibre" (RFoF) saturation through RFI. After passing two separate RF distributors for each polarization, which provide separate channels for each band, the signals are fed into eight Up/Down converters (UDC). Up/Down conversion means that all signals are first brought up to 20 - 22 GHz. After a band pass filter (BPF) for out-of-band rejection, a mixer with a 22.5 GHz fixed frequency local oscillator brings the signals down to 0.5 to 2.5 GHz. Further filtering reduces each band and polarization to approximately 512 to 1024 MHz (Niell et al., 2018). Other than in the S/X developments, the subsequent

[6] https://www.cosmicmicrowavetechnology.com/quad-ridge-flared-horns

sampling by four Reconfigurable Open Architecture Computing Hardware (ROACH) digital back-ends (RDBE) is done first, before the sampled signals are channelized into 16 channels of 32 MHz bandwidth each in a digital polyphase filter. Time tagging also takes place in the RDBE.

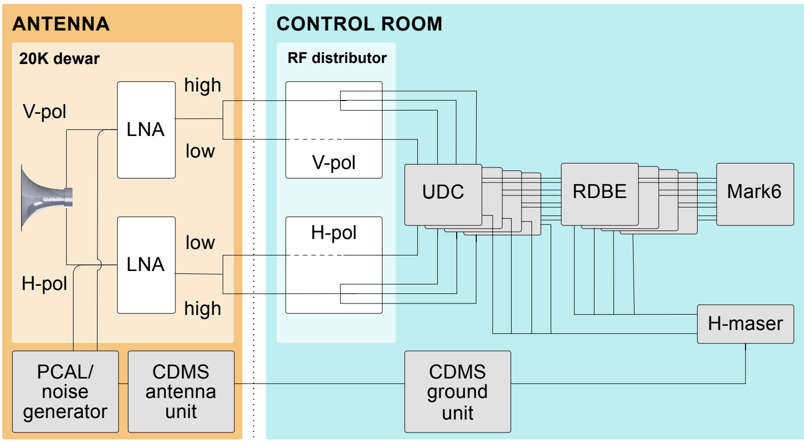


Figure 6.4: Data flow of VGOS system (from Niell et al. 2018), LNA = Low noise amplifier, UDC = Up/Down converter, RDBE = Reconfigurable Open Architecture Computing Hardware (ROACH) digital backend.

In the AUSTRAL telescopes, no separation is made and all bands are transferred by the same RFoF system. Having all the VGOS bands sent via RFoF has some significant advantages in reducing inter-band delays, unfortunately not entirely (Fig. 8.1). Digitization and channelization is done with a DBBC3 (Digital BaseBand Converter).

At the Wettzell VGOS Twin Telescopes, Up/Down converters are located next to the LNAs in the feed cone of the telescopes. They produce Intermediate Frequencies (IFs) in the range of 500 to 1500 MHz for each band directly. These are then transferred to the control room by eight coaxial cables (4 bands x 2 polarizations).

In all these configurations, the recording hardware may be Mark 6 units or FlexBuff RAID systems. At a later stage, the FlexBuff data is always transferred via the Internet while Mark 6 units may still be shipped physically or may be used as interim storage as well. With four bands of 1 GHz bandwidth each, two polarizations and a 50% recording cycle, the total volume per telescope is approximately 86 TB.

6.4. Special *Level 1 data analysis* considerations

To process the recording of signal streams from two linear polarizations, designated as X and Y at both telescopes requires the correlation of four products, $X_{(1)}X_{(2)}$, $X_{(1)}Y_{(2)}$, $Y_{(1)}X_{(2)}$, and $Y_{(1)}Y_{(2)}$, and to implement the proper combination of the products (Appendix C). The reason is that it is only with great technical effort and expense that the two linear polarization planes are kept parallel for all telescopes in a network during the observations. The combination of all products is performed with the visibilities of all cross products resulting in a single array of visibilities which represent the total intensity distribution of quasi polarization-free observations.

Since the signals of each band and each sub-band pass through separate hardware components, phase distortions are inevitable. For this reason, a proper and reliable phase calibration system (Sec. 7.1.1) is essential for the VGOS goals to be met. At present, remaining random phase biases among polarizations and sub-bands are a great obstacle still to be solved.

The existing chain of *Level 1 data analysis* handling of S/X band observations has reached a solid level of maturity. However, for the determination of delay and phase observables from VGOS observations there is still room for further development. The two areas are (1) the combination of the multi-polarization products and (2) the joint estimation of the delay and the total electron difference. Please see Appendix C on polarization issues and pre-fringe-fitting data handling.

The wide bandwidth brings with it the need for addressing two frequency-related issues. The first one is dispersion caused by all charged particles between the radio sources and the telescopes. These have quite a large, though known, frequency-dependent impact (Sec. 13.5.2). The second one is the effect of source structure on the visibilities and, thus, the delay observable. The latter is not so well understood because maps of the brightness distribution at the individual VGOS bands are not yet available. However, by now source structure effects are assumed to have as large a negative impact on VGOS observations as measurement noise and the wet atmosphere (Anderson and Xu, 2018).

The dispersive effect of the ionosphere and the wide range of frequencies (2 - 14 GHz) inhibit the assumption of a single ionosphere delay as it is applied individually in the fringe fitting process of S and X band observations. To the contrary, the joint delay, phase, and delay rate determination in the fringe fitting process needs to take into account the non-linear frequency-dependent ionospheric delay effect.

The joint estimation of the delay and the total electron content difference between two telescopes in the fringe fitting process adds another degree of freedom to the data. While in the initial processing of the individual bands, the fringe amplitude is maximized for the single band delay, the phase delay rate, and the group delay, VGOS processing requires that the differential total electron content (dTEC) be added as a fourth dimension for the maximization criterion of the fringe amplitude.

The basis of the following considerations are Eq. 13.68 for the impact of the ionosphere on the phase observable, which goes as -constant/ν, and Eq. 13.70 for the group delays, which follows +constant/ν^2. The signs are important, negative for the phase, which means that the phase is accelerated, and positive for the delay, meaning a retardation.

First, to get an impression of the magnitudes of the ionosphere affecting the group delays and the phases in each band, Tab. 6.2 lists the ionospheric phase and delay impacts for dTEC = 3 TECU according to Eqs. 13.68 and 13.70.

Table 6.2: Impact of ionosphere for dTEC = 3 TECU in different frequency bands (first two lines each are current lowest and highest allocations according to Tab. 6.1 with the third line indicating highest frequencies when 1 GHz bands will be populated).

Band	Upper bandedge MHz	Group delay ps	Approx. slope ps/100 MHz	Phase rad	Approx. slope rad/100 Mhz
A	3032.40	438.8		-8.35	
	3480.40	333.0		-7.28	
	4056.40	245.1	-18.9	-6.25	0.205
B	5272.40	145.1		-4.80	
	5720.40	123.2		-4.43	
	6296.40	101.7	-4.24	-4.02	0.076
C	6392.40	98.7		-3.96	
	6840.40	86.2		-3.70	
	7416.40	73.3	-2.48	-3.42	0.053
D	10232.40	38.5		-2.48	
	10680.40	35.3		-2.37	
	11256.40	31.8	-0.65	-2.24	0.023
E	12472.40	25.9		-2.03	
	12920.40	24.1		-1.96	
	13496.40	22.1	-0.37	-1.88	0.015

With $d\phi/d\nu = 2\pi\tau$, a positive group delay is related to a linearly increasing phase. In Fig. 6.5, the non-dispersive phase behavior of the group delay (green line with diamonds) and the dispersive phase behavior of the ionosphere retardation (blue line with circles), together with their sum (magenta line) are displayed for a differential total electron content (dTEC) of 3 TECU and a group delay of 300 ps. The sum of the two effects as displayed here in magenta is a generic example of the phase behavior which will be hidden in the visibility data before the group delay observable will be composed in the fringe fitting process. One of the main tasks there then will be to separate the ionosphere effect and the group delay.

In *fourfit*, the fringe fit is performed by finding that group-delay which maximizes the coherent sum of the residual fringe phasors over time and frequency. When the ionospheric dispersion is non-negligible, *fourfit* is told to search "over a grid of potential differential ionosphere TEC values as well to find a maximum coherent phasor sum via a parabolic interpolation of the gridded values. This method of finding the maximum coherent sum w.r.t. the group-delay has been shown to be equivalent to using least-squares estimation in the region of the maximum" (Cappallo, 2015). Through an empirical test for the frequency setup above (Tab. 6.1), Cappallo (2015) showed that the correlation coefficients between the estimated group delay and dTEC value can be as large as 0.93. In addition, the correlation coefficients between the reference phase ϕ_0 and dTEC are as high as 0.99. Although *fourfit* rather works along the lines quoted above, the least squares test shows that the separation of dTEC from the group delay still leaves some room for improvement. Unfortunately, an attempt to reduce the correlations through a frequency selection different to that of Tab. 6.1 fails because the partial derivatives of τ $(\nu_i - \nu_0)$, ϕ (1), and dTEC $(-1.3445/\nu_i)$ all remain in the same regime and, thus, cannot vary sufficiently to achieve a noticeable degree of de-correlation.

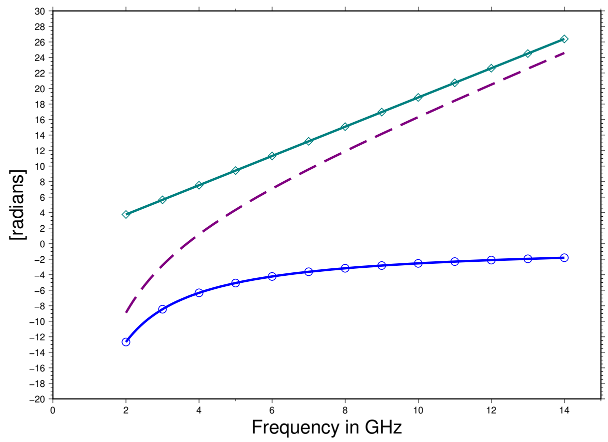


Figure 6.5: Phase versus frequency for dTEC = 3 TECU and a group delay of 300 ps. Green line with diamonds = non-dispersive delay, blue line with circles = dispersive ionosphere acceleration, dashed magenta line = sum of dispersive and non-dispersive effects on the delay

.

Kondo and Takefuji (2016) propose a different concept for the delay, phase, and dTEC determination. Prerequisite in the VGOS context is that data from two linear polarization signal chains are already combined to "polarization-free" total intensities as is the case in the (pseudo) Stokes I approaches. The authors use a reference scan of a strong source to identify non-linearity in the phases, including instrumental and dispersive (ionosphere) effects across all bands and sub-bands first. In a second step, the corrections are applied to the visibilities of all other observations as well, and interim delays are determined which still contain the dispersive ionosphere effects. In the last step, the dTEC values are determined starting with a coarse estimation of dTEC with a group delay model in order to avoid phase ambiguity problems. The subsequent precise estimation then works on the basis of the phases void of any ambiguities. While the first approach had a poor yield for weak sources, a new method provided better results (Kondo and Takefuji, 2019). Here, a dTEC search function with a (default) search range of ±20 TECU was introduced prior to the least squares estimation.

7. Special elements of observing systems

7.1. System calibrations

From the feed horn of the telescope, the signals pass through various components of the receiving system before they are digitized and time tagged (see Sec. 3). It is quite natural that the signals suffer from system delays and phase changes which are neither constant nor predictable (Clark et al., 1985). To calibrate these changes, two electronic components have been devised which are now part of the standard VLBI hardware of most radio telescopes. The first is the phase calibration (often abbreviated as phasecal or pcal) system, the second is the cable calibration system (often abbreviated as cablecal or delaycal). The systems together serve to calibrate instrumental delay and phase changes. The cablecal system consists of a ground unit and an antenna unit (Fig. 7.1). In addition, many telescopes utilize noise calibration systems. These are employed to determine the thermal-equivalent of the system noise, which is called *noise temperature* and is measured in Kelvin.

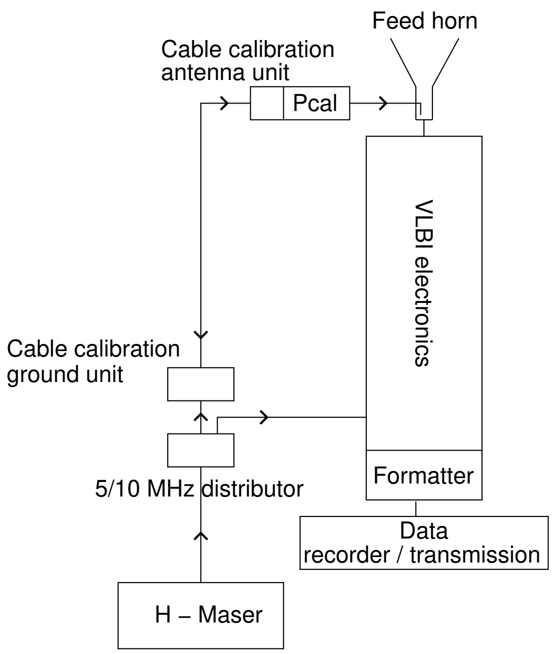


Figure 7.1: Equipment for phase and cable calibration.

7.1.1. Phase calibration system

In the phasecal system, a comb of Dirac pulses of 1 ps (legacy S/X system) or 5 or 10 ps (VGOS system) repetition rate (Fig. 7.2) are injected near the feed horn (Fig. 7.1). Originating in an implicit Fourier transform from the time into the frequency domain, the pulses appear as spikes repeating every 1, 5 or 10 MHz, respectively, over the whole radio frequency spectrum, which means that they appear in all of the frequency sub-bands. Since these tones possess a phase, they can be used as phase calibration signals. For this purpose they are added to the noise from the quasars, travel through the same electronic components, and are finally recorded as part of the combined signal. The phase calibration tones are extracted again as part of the correlation process (Sec. 10). While the phase-cal tones need to be strong enough for good detection, the power in the tones should be kept at about 1% of the quasar noise to avoid reducing the quasar signal-to-noise ratio (See Fig. 5.14 for a real analog bandpass where phasecal spikes are clearly visible every 1 MHz.). In VGOS systems with the total bandwidth spanning 12 GHz, the amplitude tends to decrease towards the higher frequencies due to the finite rise time of the Dirac pulses. For this reason, the pulse generation components should be designed to allow for extremely sharp pulses.

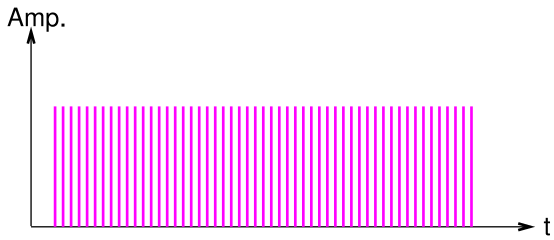


Figure 7.2: Dirac pulses in the time domain

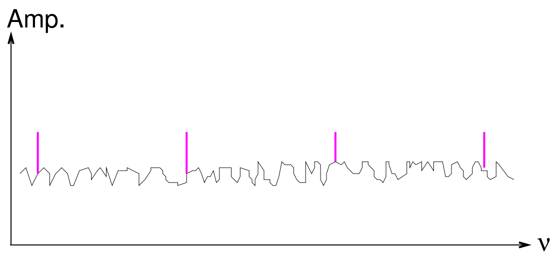


Figure 7.3: Phase calibration tones in the frequency domain within a single sub-band of, e.g., 16 MHz width.

Phase calibration serves two purposes. The first one is that inherently the phase calibration marks the time epoch when a certain signal component has actually arrived at the feed horn. This can be considered as a timing reference. The second purpose is that for bandwidth synthesis, multiple sub-bands are processed in the system, i.e., digitized and down-converted, and dispersive effects occur to the phase of each sub-band. These lead to spurious delays according to Eq. 2.6 which are variable. The phase calibration information is extracted during correlation (Sec. 10.6) and applied in the fringe fitting process (Sec. 11.4).

7.1.2. Mark III phase and delay calibration system

The unit producing the phase calibration signal is located very close to the feed horn (Fig. 7.1). It receives its reference frequency from the H-Maser through a 5 MHz distributor which also supplies the local oscillators with phase coherent frequencies. This link (normally between the electronic rack in the control room and the receiver unit near the feed horn) is subject to changes in the electrical path caused primarily by temperature variations and by bending and twisting of cables. Coefficients of phase delay sensitivity to temperature of coaxial cables can reach up to 100 ppm/K[7], while optical fiber has a sensitivity of less than 10 ppm/K. For a 30 m cable, this amounts to 10 ps/K and 1 ps/K, respectively.

Since the length change causes an equivalent change in delay and the temperature-driven cable delay changes on time scales of minutes to hours, a calibration system was designed and implemented in the early Mark III system era (Rogers, 1980) as a component of the phase and group delay calibration system. It is called cable calibration system, a.k.a. delay calibration system, and was realized as a two-way system with a ground unit and an antenna unit (Fig. 7.2). In the Mark III system, a 5 MHz signal is used to generate the phase cal tones as described above. For this purpose and for calibrating the variations of the up-link, the up-link signal is offset in frequency by 5 kHz by the calibration ground unit. In the antenna calibration unit, the signal is divided into input for the calibration signal and an input for a reflection down to the ground unit. The 5.000-MHz component of the first of these two signals is the input to the phase calibration signal generation unit. Here the rectangular flanks of the signal are fed into a tunnel diode for generating 1 MHz pulses from every fifth positive pulse.

The second signal is down-converted to 4.995 MHz and sent back to the ground unit, where the 5 kHz (negative) frequency offset is removed, and where the phase is compared to the up-going 5 Mhz signal (Fig. 7.4). More concretely, the signals whose phases are to be compared are mixed down from 5 MHz to 25 Hz using a local oscillator signal with frequency 5 MHz minus 25 Hz (Rogers, 1980). Through this manipulation, the 5 MHz signal is spread in time by a factor of $200,000$ ($\frac{5MHz}{25Hz}$). The outputs of the comparator are two 25 Hz square waves, one from the antenna unit and a reference signal generated from the 5 MHz signal from the H-Maser. The relative phases of these two 25 Hz signals are equal to the relative phases of the two 5 MHz input signals. For the difference of the outgoing and incoming trigger points at the ground unit, each count of an off-the-shelve start-stop counter operating at a 5 MHz clock rate is, thus, equivalent to 1 ps ($= \frac{1}{5MHz}/\frac{5MHz}{25Hz} = \frac{1}{(5MHz)}/200,000$) or 1.8×10^{-3} degrees (Rogers, 1980).

Since the counter readings are differences of round-trip delay, the one way delay or phase difference is half of that. In addition, the counter readings monitor only the variations, and, thus, they are only relative to some arbitrary reference, e.g., the phase at the start of the session. Furthermore, the counter cannot distinguish whether the trigger signal of the returning wave comes before or after the trigger signal of the outgoing wave (25 Hz seems to be large enough for this but the initial state of the counter prohibits a clear distinction). For this reason, the counter reading can have either a positive or a negative meaning, although the reading itself is always positive. To figure out the sign, station staff are asked

[7] https://ivscc.gsfc.nasa.gov/meetings/v2c_wm1/phase_stability.pdf

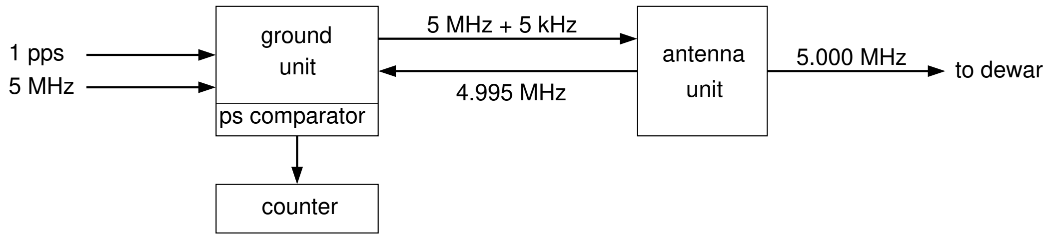


Figure 7.4: Legacy Mark III cable calibration system

to carry out a verification measurement. At the beginning and the end of a session, apiece of cable of approximately 30 cm length is inserted in the uplink cable and a cable reading is performed. If this reading is larger than the reading without the extra cable length, the sign of the cable calibration recordings is positive; if it is smaller, the sign is negative.

All readings are recorded in the station log to be used for calibration in the data analysis process (Sec. 13). Often, the effect of the extra piece of cable inserted at the beginning and/or the end are clearly visible when the delay-cal values are plotted.

Arthur Niell: "However, the initial purpose of the cable-cal calibrator was to monitor the long-term stability of the signal chain. Each calibrator cable was carefully measured with an accuracy of a few picoseconds before being sent to the station. If monitored by the station, any change in value could provide either an alert to unexpected changes or a record of changes in the signal chain."

7.1.3. VGOS Cable Delay Measurement System (CDMS)

For VGOS systems, MIT Haystack Observatory developed a new calibration system for determining instrumental effects such as cable delays, signal phase, and signal amplitude. The operation of the cable length calibration is similar to that of the Mark III system (Sec. 7.1.2) with cable delay and phase calibration facilities realized within a ground and an antenna unit. The main differences are that the phase calibration pulses are separated by 5 μs producing tones every 5 MHz, and the lack of a counter for determining the cable delay variations.

The ground unit requires externally provided signals of 10 MHz, 5 MHz, and 1 pps with the stability of a hydrogen maser. The delay stability is designed to exceed 1.8×10^{-14} at 30 s or 1×10^{-16} at 50 min (Allan standard deviation).

The CDMS also contains an amplitude calibration feature in the form of a broadband noise source. The noise amplitude is calibrated using a programmable controller for the noise signal, which is injected at the antenna feed with a 50% duty cycle square wave with a switching rate of 5 -– 100 Hz. This noise signal can be detected by a digital backend to provide a time tagged system temperature table that is delivered with the primary fringe visibility data.

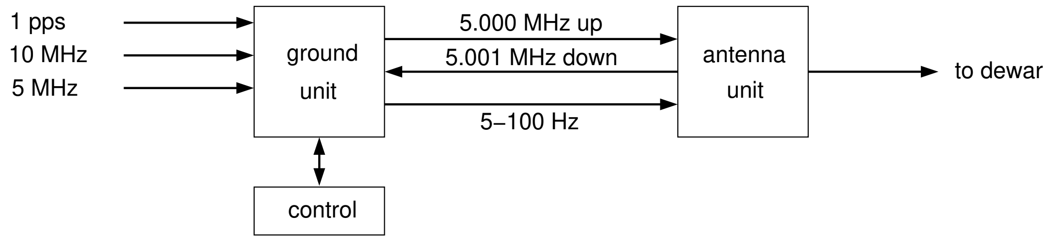


Figure 7.5: MIT Haystack Observatory cable delay measurement system (CDMS). Here, the measurement signal on the downlink is 5 MHz plus 1 kHz.

7.1.4. VGOS proxy cable calibration

Unfortunately, not all telescopes possess a delay calibration unit with sufficient performance, and on some occasions the unit may malfunction. For these situations, a proxy for the in-situ measurements has been devised to be employed at the analysis stage. The so-called proxy cable delay measurements are based on the idea that any electrical path length variations in the 5 MHz cable are common to all bands, sub-bands and polarizations, and thus the delay should show up as a linear phase change vs. frequency which contributes to the phase-cal in addition to any variation due to the signal chain instrumentation.

The cable-cal delay usually varies slowly with time since the primary cause for variation is changing temperature of the cable. By slowly is meant that over the length of the piece-wise linear clock approximation in the geodetic estimation, the cable delay change can also be approximated as linear and this will be subsumed by the clock parameter.

In addition to providing information on the slow changes in cable delay, the cable-cal measures variations due to bending and twisting of the cable caused by antenna motion in azimuth and elevation. For VGOS this timescale is a minute or less, and due to the scheduling requirement to cover the full visible sky at a site as quickly as possible, all azimuths and elevations will be sampled in about ten minutes. If there is a variation in the cable delay that is related systematically to antenna orientation, this will dominate the variation within the clock estimation interval (after removal of the linear-with-time clock term) and can result in an apparent antenna position displacement. Since the phase cal across all frequencies provides a measure of the electrical path length from the origin of the 5 MHz reference, up the cable to the phase-cal generator, and down through the signal chain, the delay calculated from all of the phase cal tones within a scan can be used as a measure of the cable delay, similar to what would be obtained by the CDMS. In particular, this proxy cable-cal responds to the orientation-dependent delay changes in the 5 MHz cable and can be applied as though it were a CDMS (Niell et al., 2018).

7.2. Observing frequency predefinition

Many people ask why the sky frequencies of S/X were initially chosen to end with .99 MHz and why the VGOS frequency sub-bands all end with .40 MHz. The reason, as with many other decisions, lies in the circumstances and technical feasibility when the original system had been developed. In this respect, the predefinition of the .99 frequencies goes back to the late 1970ies when the original Mark III system had been developed with the available hardware for 2 MHz wide sub-bands/channels.

At that time, the phase calibration signal was chosen to have a pulse repetition rate of $1 \times 10^6/s = 1$ MHz resulting in frequency tones at full MHz intervals at the front end of the signal chain. With a selection of ".99", a sub-band/channel at sky frequency would range, e.g., from 8210.99 MHz to 8212.99 MHz with the phasecal tones at 8211.00 MHz and 8212.00 MHz. After two steps of down-conversion with a total local oscillator frequency of 8210.99 MHz, baseband is from 0 to 2 MHz and the phasecal tones are located at 10 kHz and 1010 kHz.

As a brief sideline, it should be mentioned that the MkIII correlator had no way to extract the 1010 kHz tone. Once the MkIV processor came along, 1010 kHz could be extracted (along with 2010, 3010,... kHz as the channel bandwidths got wider). This had the advantage that these tones could be used, should 10 kHz be affected by spurious signals. In addition, it opened the path for the development of multitone phase calibration (Sec. 10.6 and Sec. 11.4).

The first 10 kHz frequency, which is the result of the frequency setup scheme, is chosen on purpose as Brian Corey (priv. comm.) explains: "Unlike the phasecal extractors in the MkIV and more modern (e.g., DiFX) correlators, the MkIII extractor could process only tone frequencies for which there was an integer number of samples in one quarter of a period. Hence, for an analog baseband bandwidth of 2 MHz (sample rate 4 MSps), 10 kHz was a permissible frequency (100 μs period –> 100 samples / quarter period). But a frequency like 400 kHz was not permissible (2.5 μs period –> 2.5 samples / quarter period).

10 kHz was chosen over nine other possible frequencies because it had the smallest non-static phase error[8]. It is "non-static" because the phase error changes as the phase changes. In the case of 10 kHz, e.g., the largest non-static error is caused by the tone that, for a 2-MHz BW, was at 2010 kHz in the analog domain, but that got folded over to 1990 kHz after sampling (If the bandpass filters cut off sufficiently sharply at the upper end of 2 MHz, then the 2010 kHz tone would not be problem. But the filters are nowhere near good enough for that.). Because the 2-level phasecal extraction waveform in the MkIII processor has nonzero harmonic content at every odd multiple of the fundamental tone frequency, the aliased tone at 1990 kHz (199th harmonic of 10 kHz) will contribute to the signal extracted for 10 kHz. If the LO phase changes, the phase of the 10 and 2010 kHz analog tones will shift by the same amount, but the phase shift of the 10 and aliased 2010 kHz digital tones will have opposite sign, and the error in the extracted phase will change. This non-static error is more serious than a static error that is independent of LO phase.

Phase errors of this sort with the MkIV and modern digital correlators are smaller than those in the 1979/1992 memo because they use extractors with more than just two levels.

[8]https://library.nrao.edu/public/memos/vlba/test/VLBAT_29.pdf and 1979 TN embedded

Plus, they are much less restrictive in terms of what tone frequencies they can handle.

With a sky frequency of x.99 MHz and with phasecal tones generated at y.00 MHz, the baseband phasecal frequency closest to baseband DC is 10 kHz, as desired for minimum non-static error in MkIII days.

Note that, while spurious signals are a major and persistent source of phasecal error, they do not play a role in setting the '.99' part of the sky frequencies. They definitely must be considered when selecting which tones within a channel to use to calculate phasecal phase (and delay). But the amplitude and phase of a spurious signal don't depend on the sub-MHz portion of an LO setting (unless the spurious signal originates after the frequency conversion, which is rarely the case). Rather, they depend only on the tone frequency."

The VGOS systems use LSB frequency channels and the local oscillator in the Haystack-style up/down converter is restricted to frequencies that are integer multiple of 400 kHz. For channel (center) frequencies, there are thus only five choices for the sub-MHz part (0.0, 0.4, 0.8, 0.2 (1.2 - 1 MHz), and 0.6 MHz (1.6 - 1 MHz)), and 0.4 MHz is as good as any (and much better than 0.0, which would have many problems with interferences). Because the VGOS channels are lower sideband and the VGOS pcal tones are spaced 5 MHz apart at most sites, the first (lowest-frequency) baseband pcal tone can therefore be at 0.4, 1.4, 2.4, 3.4, or 4.4 MHz, depending on the value of the full MHz setting. (Brian Corey, priv. comm.)

In short: "The 0.40 MHz offset in the VGOS broadband system is the smallest offset that can be obtained in the UpDown Converter (UDC) due to multiplication of the reference signal to generate the local oscillators in the UDC. This is specific to the Haystack signal chain design and cannot be reduced." (Arthur Niell, priv. comm.)

8. Mixed mode operations

For some time, legacy S/X radio telescopes and VGOS infrastructure will continue to exist and operate in parallel. In order to allow for a link of the reference frames, both terrestrial and celestial, which would otherwise stay independent next to each other, operations of the two types of telescopes in joint observing sessions, called mixed mode sessions, is of great interest to the community. Although this may sound trivial in general, the complications lie in the details and especially in the different hardware concepts at the various telescopes.

For a start, let's recall that the front and back-ends of the legacy telescopes normally cover the frequency ranges from 2200 to 2400 MHz and 8200 to 9000 MHz while the VGOS frequency bands are currently defined as 3000-4000, 5240-6240, 6360-7360, and 10200-11200 MHz. Per se, there is thus no overlap in frequency ranges, and legacy front- and back-ends are rather inflexible for changes in receiving frequency ranges far away from those for which the equipment is optimized. For this reason and because front-ends should cover the full range from 2 to 14 GHz continuously, the VGOS receiving equipment has to be adapted for the purpose of mixed mode observations.

The first issue is the characteristics of the feed horn. For example, the Quad Ridge Flared horns (QRFH) have a nominal operating range of ≈2.3 - 14 GHz with a slow performance roll off below 3 GHz. In this respect, the Eleven Feeds are better because their nominal range starts at 1.4 GHz and frequencies around 2200 to 2400 MHz produce at least if not more power than above 3000 MHz.

The most important peculiarities, however, are related to the down-link scheme. The Haystack developments in the current prototype system split the LNA output into a low (below 5.0 GHz) and high frequency (above 5.0 GHz) regime (Fig. 6.4) and transfer this to the control room by coax and fiber optics cables, respectively (Niell et al., 2018). The "high" band is realized by a 5 GHz high pass filter while the 3 GHz signals are sent via co-axial cable to avoid potential RFoF saturation through RFI. Consequently, S band is transferred in this "low" chain. A difficulty with mixed-mode observations comes from the sampler in use. While the DBBC3 can place baseband channels freely within the input 4 GHz band, the RDBE/RDBE2 systems of the Haystack developments use a polyphase filter bank with fixed channel spacings. This limits the selectable frequency channels, which do not necessarily match the standard S/X setups, and then requires a modification of the legacy frequency spacing.

The AUSTRAL front-end design with QRFH feed horns splits the output of the LNAs with one side going through a 3 GHz high-pass filter before entering the RF over Fibre (RFoF) system. The other, unfiltered, output is connected into a legacy S band back-end with a fixed 1900 MHz oscillator and bandpass filters. This is sent over existing coaxial cables to the control room. So, it is this existing channel with a remotely-controllable RF switch (red in Fig. 8.1) which permits to easily toggle between observing in the S/X and VGOS modes.

At Wettzell, the Ws telescope is equipped with additional hardware for decoupling S and

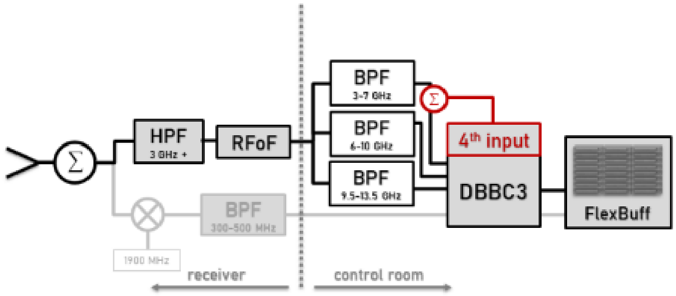


Figure 8.1: Flow chart of Austral mixed mode setup (adapted from McCallum et al. 2022). HPF = high pass filter, BPF = band pass filter

X band signals from the broadband LNAs with subsequent independent channel design. An important feature of the hardware is a 90° hybrid linear to RCP circular polarization element. Directly in the feed cone, the output of this is down-converted to the 500 to 1500 MHz IF band and transferred to the control room and the DBBC3 by separate coax cables for the RCP S and X signals.

Except of the Wettzell design, which produces data, which is directly fit for a conventional RCP data correlation, all other VGOS telescopes produce separate horizontally and vertically polarized data streams. This, thus, requires legacy RCP vs. horizontal and legacy RCP vs. vertical polarization correlations and subsequently a combination with mixed mode fringe fitting (McCallum et al., 2022).

9. Scheduling process

Observing VLBI sessions requires an active control process because it has to be guaranteed that the telescopes forming one or several baselines point at the same quasar for the same period of time. In addition, recording of the data has to be synchronized as well since registration is not continuous to make optimal use of the recording media. Under these premises, observing schedules are prepared a few weeks before the observing date. They contain the start and stop times for every telescope in the network for every individual observation. Assuming a network of up to 20 telescopes, multiple configurations of subsets of these telescopes form so-called scans of one quasar at a time. The composition of these subsets always depends on the visibility of the quasar from a certain area on the Earth's surface. In the end, a schedule for an observing session of 24 hours may consist of thousands of scans. The word scan is actually used if two or more telescopes simultaneously observe the same radio source. Each scan produces $n \cdot (n-1)/2$ individual delay observations for n telescopes in the same scan.

The preparation of the observing schedules is of great importance to the overall results because it defines the geometric configuration of the parameter estimation. In this respect it is comparable to geodetic network optimization (e.g., Grafarend and Sansó, 1984, Amiri-Simkooei et al., 2012).

The planning of VLBI observations, in fact, is rather complicated because there are many parameters which have to be taken into account. The first question to answer always is whether a quasar is actually above the horizon at the telescope. This can be deduced from a pure geometric consideration resulting from the position of the quasar in a sky-fixed celestial reference frame (see Sec. 13.1.2) and the coordinates of the telescope in the terrestrial frame transformed into its instantaneous location in an geocentric celestial reference frame. To first order this is just a simple rotation about the Earth rotation axis with the local hour angle h_{loc}. With some generalization it is computed with the right ascension α, the longitude of the telescope λ, Universal Time UT, and Greenwich Mean Sidereal Time (GMST) at 0h UT according to

$$h_{loc} = \mathrm{GMST_{0h}} + \mathrm{UT} \cdot 1.00274 + \lambda - \alpha. \tag{9.1}$$

For any radio source with its right ascension α and declination δ and any telescope with its geographic latitude Φ and its instantaneous hour angle the respective azimuth A and elevation ε can be computed according to Mueller (1969)

$$\tan \mathrm{A} = \frac{-\cos\delta \sin h_{loc}}{\sin\delta\cos\Phi - \cos\delta\cos h_{loc}\sin\Phi} \tag{9.2}$$

$$\sin\varepsilon = \sin\Phi\sin\delta + \cos\Phi\cos\delta\cos h_{loc}. \tag{9.3}$$

Querying the local horizon mask of each telescope, which may be limiting the observable section of the sky, the question is answered whether a radio source is observable at a certain instant of time. By default, VLBI observations are carried out down to the threshold of $\varepsilon \geq 3°$.

Due to the fact that cable links between the fixed ground part and the turning elements need to be secured from over-twist, all azimuth-elevation telescopes are limited in their azimuth rotations. Normally, a full circle is augmented by some reasonable arc length in clockwise and counter-clockwise directions to allow for continuous tracking at all directions. Technically, this is often realized by a spiral tray holding the cables and allowing the diameter to change depending on the rotations of the telescope (Fig. 9.1). At the Wettzell 13.2 m twin telescopes, the wrapping mechanism keeps the total curvature of the twist constant over the whole cable lengths but the maximum bending is shifted depending on the azimuth.

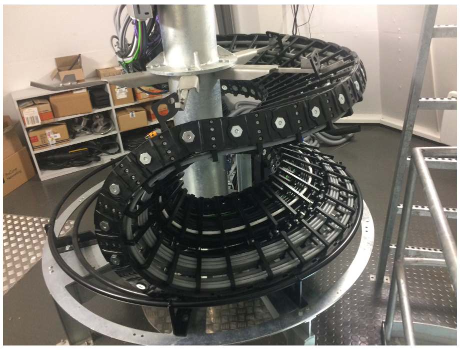


Figure 9.1: Cable wrap mechanism below the turning heads of the Wettzell 20 m telescope (left) and the 13 m telescope (right).

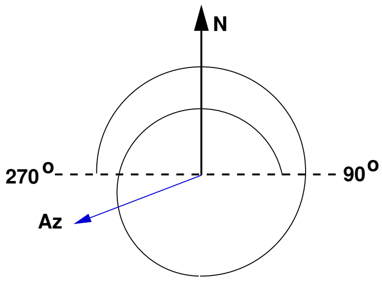


Figure 9.2: Example of cable wrap limitations (-90° to +90°). N.B.: Position **Az** may be reached by two different azimuth values

In the scheduling process, special care has to be taken that the duration of the slew from one azimuth to another is computed correctly taking into account that sometimes the shortest path is blocked by the antenna limits. These limits are tabulated, for example in the SKED catalog file, in clock-wise direction. In the example in Fig. 9.2, the left-most limit is -90° and the right-most, after a full clock-wise turn, is +90°. To keep the limit count always positive, SKED would list these limits as 270° and 810°, respectively.

For each source the duration of the observation has to be calculated. This mainly depends on a predefined minimal signal-to-noise ratio (SNR) which is normally set at a level of 25 to 35. For each source, the necessary integration time is computed according to Eq. 3.3 depending on the correlated flux density of the source and the sensitivity of the telescopes represented in system equivalent flux densities (SEFD) according to Eq. 3.13. For the damping caused by the atmosphere the SEFD at zenith (S_{sys}^{zenith}) has to be scaled according to Gipson (2016) p. 170 by

$$S_{sys}(\varepsilon) = S_{sys}^{zenith} \cdot f(\varepsilon). \tag{9.4}$$

In its simplest form, the scaling factor $f(\varepsilon)$ is a mapping function just as $1/\sin\varepsilon$. A more complex model

$$f(\varepsilon) = \sum_{i=1}^{n} \frac{c_i}{(\sin\varepsilon^y)^i} \tag{9.5}$$

is being used in the SKED scheduling program (Gipson, 2016), where n is the number of terms (usually 1 - 3), c_i is the coefficient for the ith term, and y is the power of the $\sin\varepsilon$ term $(0 \leq y \leq 1)$. Since it takes quite some effort to determine the coefficients and power law, these are available only for a few radio telescopes such as those of the Very Long Baseline Array (VLBA) (Petrov et al., 2009) with c_1 close to 1.

These mapping functions lead to an increase of the SEFDs already by a factor of 2 at $\varepsilon = 30°$ increasing sharply below. For this reason, observations close to the horizon need much longer integration times than those near zenith.

The final observing schedule is constructed in a sequential forward mode where new observations are being selected to fulfill an optimization criterion. Most operational observing schedules are generated to optimize the sky coverage at each telescope in a way that the hemisphere above each telescope is filled homogeneously with observations as much as possible for a good estimation of the atmospheric refraction parameters (Fig. 9.3 and 9.4).

It should be noted here that due to the fact that the atmospheric refraction parameters are often estimated as linear splines with interval lengths of, e.g., 1 hr, the sky coverage is actually evaluated only for the hour preceding the current instant.

The generation of the observing schedules is a contemporary field of research and many innovative optimization schemes have been devised (e.g., Baver et al. (2012), Steufmehl (1994), Sun et al. (2014), and Uunila et al. (2012)). In the scheduling process itself, we assume a start scenario where all telescopes point at the same quasar. The question to answer is which is the next observation fulfilling the optimization criterion. To answer this,

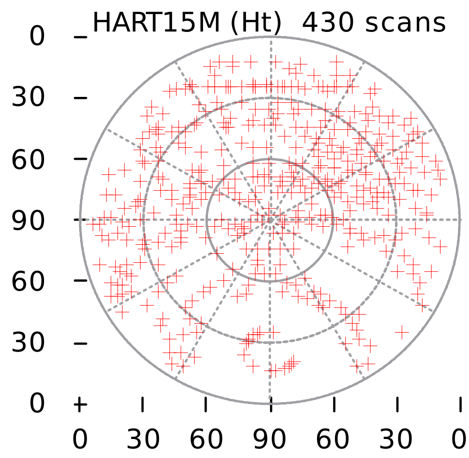


Figure 9.3: Local sky coverage of observations at telescope HART15 (South Africa).

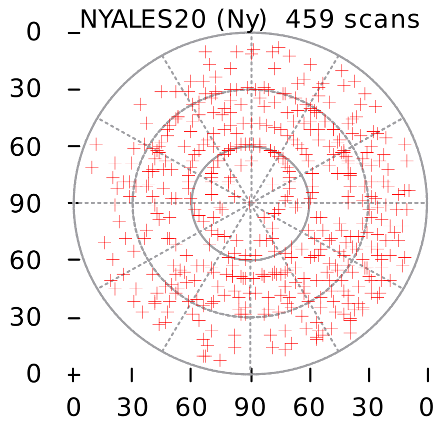


Figure 9.4: Local sky coverage of observations at telescope NYALES20 (Svalbard, Norway).

for all possible candidates the next possible start time is computed from the slew rates of the telescopes and the minimum scan lengths for each baseline according to the SNR threshold. The final selection is then taken in agreement with some predefined minor options such as maximum scan length, minimum separation between observations of the same source. For more details of the geodetic scheduling procedure see Gipson (2016).

10. Correlation

10.1. General overview

The VLBI time delay τ is defined as the travel time of a plane radio wave which passes station 1 and then reaches station 2. At this point we are not going into detail how the reference point at each station is defined, but rather focus on the question how we can compare the recorded bit-streams from each station and extract quantities like group delay, phase delay, delay rate and signal-to-noise ration (SNR) , respectively correlation amplitude. Before we have a look into theoretical aspects and study the actual correlation process, it might be good to reflect on the geometrical situation and the technical aspects of VLBI.

If we consider two VLBI stations distributed arbitrarily on the globe, we can easily identify the range of possible delays to be within

$$\frac{-R_e}{c} \leq \tau \leq \frac{R_e}{c}, \tag{10.1}$$

where R_e is the radius of the Earth and c denotes the speed of light. Thus, any VLBI delay has to lie between these boundaries which evaluate to approximately 21 ms. Without stating explicitly, we have assumed for this simple estimate, that the station clocks are synchronized to UTC sufficiently well. While this is not a straightforward task and requires the use of GNSS receivers, VLBI stations can normally synchronize their time-tagging units to UTC within ± 1 μs. Revisiting the VLBI overview depicted in Fig. 1.2 one has to consider that the situation shown there is a static snapshot. During an actual VLBI observation, two main effects due to temporal variations need to be considered. First, the Earth is rotating during a scan is recorded. Thus, similar to GNSS, one needs to take into account a Doppler effect which impacts the data stream at each station in a way that the recordings suffer from delay rate effects. As similar but smaller contribution might also come from the frequency standard at each station, which very likely drifts from its nominal frequency and thus adds another delay rate-like effect. Ignoring higher order temporal effects, we can thus state that the correlator's task to find the delay τ between two stations can only be achieved if, at the same time, the delay rate $\dot{\tau}$ between the recorded data streams is taken into account properly. This means, that the correlator's task can be summarized as finding the most likely values for delay and delay rate that relate the sampling data from station 1 to those of station 2.

As will be addressed in more detail later, the correlation process is actually limited to a small search time interval applying a priori delays (to be re-added again at the end) from a priori geometry and clock offset information. Furthermore, a single observation is first decomposed in data chunks of individual periods of 0.1 to 1 s, so-called accumulation periods (APs), which later on are combined again. First, each sub-band sub-band is looked at separately and at the end these are combined as well leading to bandwidth synthesis (cf. Sec. 3).

10.2. Making use of the cross-correlation theorem

Ignoring for a moment that recorded signals suffer also from non-zero delay rate effects, and assuming that signals are captured with infinite temporal and spectral resolution we can write the cross-correlation function[9] as

$$C_{12}(\tau) = \mathrm{Corr}(s_1, s_2) = \frac{1}{T}\int_{-\infty}^{\infty} s_1(t)s_2(t-\tau)dt. \tag{10.2}$$

where s_1 and s_2 denote the continuous recordings at station 1 and station 2 (Thompson et al., 2007). Since we can only sample in discrete time steps and we should consider the geometrical constraints discussed before we have to collect a sufficiently large number of samples in order to be able to identify the correlation peak that corresponds to the time delay between the two recordings. We have to restrict ourselves also to a certain bandwidth B when we collect our samples which implies that we have to obey the Nyquist theorem which suggests that we should record one independent sample every $\Delta t = 1/(2B)$. Thus, the finite integral from Eq. 10.2 turns into

$$C_{12}[\tau] = \frac{1}{N}\sum_{k=1}^{N} s_1[k]s_2[k-\tau], \tag{10.3}$$

where τ is considered to be a discrete (or integer) delay that has a granularity of Δt. Before we discuss how we can achieve a sub-sampling delay resolution, we should look for efficient ways to compute the cross correlation function.
Considering that the integral in Eq. 10.2 looks almost like the convolution of two functions s_1 and s_2 which is defined as $\int_{-\infty}^{\infty} s_1(t)s_2(\tau - t)dt$ and expressed in the form

$$s_1 * s_2 = \mathscr{F}^{-1}\left(\mathscr{F}(s_1)\cdot\mathscr{F}(s_2)\right), \tag{10.4}$$

where $\mathscr{F}$ and $\mathscr{F}^{-1}$ are Fourier and inverse Fourier transforms.

10.3. Efficient ways to obtain the cross correlation and cross spectrum functions

In order to obtain the correlation function rather than the convolution between the two signals, we only need to flip the 2nd signal's time order, i.e. changing from $t - \tau$ to $\tau - t$. However, this corresponds to a simple multiplication of the frequency axis with -1 in the Fourier domain. Furthermore, if $S(f)$ is the Fourier transform of a signal $s(t)$ then we are looking for $S(-f)$. Fortunately, we can can make use of another very useful relation that states that $S(-f) = \overline{S(f)}$ where long bar indicates the complex conjugate operator. Thus,

[9] In radio astronomy τ is often defined as the delay w.r.t. the second station, which would change the sign in the following equations

we have found a very simple and elegant way to express the correlation function $C_{12}(\tau)$ by two Fourier transforms, a multiplication of Fourier coefficients and a single inverse Fourier transform, i.e.

$$C_{12}(\tau) = \frac{1}{T}\mathscr{F}^{-1}\left(\mathscr{F}(s_1)\cdot\overline{\mathscr{F}(s_2)}\right), \tag{10.5}$$

Since we deal with discrete and finite sampling data we can replace the continuous Fourier transformation by Fast Fourier Transform (FFT) operations and obtain the cross correlation function by

$$C_{12}[\tau] = \frac{1}{N}\mathrm{IFFT}\left(\mathrm{FFT}(s_1)\cdot\overline{\mathrm{FFT}(s_2)}\right). \tag{10.6}$$

Please observe that the cross correlation function is now implicitly defined as a complex function and its absolute value would correspond to Eq. 10.2 in case s_1 and s_2 were real-valued data streams. As we will see later, the original raw sampling data are of course real valued, but down-conversion or signal preprocessing steps can be preformed easier when dealing with complex-valued data streams. Often we are not after the cross correlation function, but try to access the so-called cross-spectrum between the two recordings,

$$C_{12}[f] = \mathrm{FFT}(s_1)\cdot\overline{\mathrm{FFT}(s_2)}, \tag{10.7}$$

which can be turned into the cross correlation function at any point by applying an inverse Fourier transform. We could have also obtained the cross spectrum from Eq. 10.3, if we apply the Fourier transform on the cross correlation function, i.e.,

$$C_{12}[f] = \mathrm{FFT}\left(\frac{1}{N}\sum_{k=1}^{N} s_1[k]s_2[k-\tau]\right). \tag{10.8}$$

Thus, we have found two ways, summarized also in Fig. 10.1, to perform the correlation

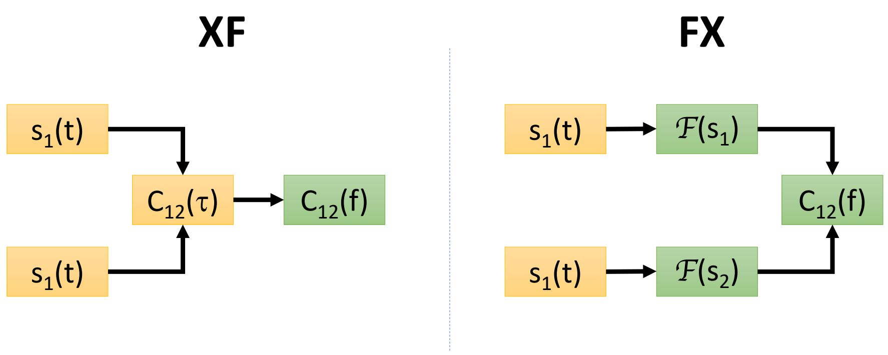


Figure 10.1: Schematic overview of XF and FX correlator architectures.

respectively the cross spectrum between the two signals s_1 and s_2. If we follow the concept of Eq. 10.8 we denote this choice as XF correlation, which indicates that the cross correlation is performed first in the time domain and then the Fourier transform is applied. If we, on the other hand, follow what is implicitly expressed in Eq. 10.7 we perform two

FFTs first and then multiply in the spectral domain, which is the reason why this choice is called FX correlation. In the end the cross-spectra will be the same, but as it turns out, FX correlation is preferable for nowadays VLBI operation due to the fact that FFT operations scale by $O(N \cdot \log_2(N))$ and thus outperform the classical XF approach already for a very small number of lags.

In order to understand why the cross spectrum appears to be the main target in correlation rather than the actual correlation function, we need to reflect on the fact that we are dealing with sampling data rates of several mega-samples per second (MSps) and thus cannot correlate the whole length of the data stream, but we will process that data in shorter pieces, so-called accumulation periods (APs), which in the ideal case even matches order-of-two FFT sizes and thus lead to a higher performance of FX type correlators. As long as we preserve coherency among those batches of spectra we can simply add them, i.e. integrate them and then perform the inverse Fourier transform on the stacked spectrum. Moreover, having access to the complex spectra in each of these batches allows us to compensate for delay and delay rate effects in a computational efficient way so that coherency can be preserved rather straightforward. Since the relation between the cross spectral characteristics and the corresponding correlation function is crucial for understanding the overall correlation process, Fig. 10.2 illustrates three simple cases that allow the reader to reflect on the concept of correlation. For the sake of simplicity we have created a short 64 sample data set of purely complex random data. From this data we created a second dataset which was circularly rotated by 5 lags. Thereafter, both data streams were Fourier transformed and the cross spectrum was obtained following Eq. 10.7. The first row in Fig. 10.2 contains three plots depicting the corresponding cross spectrum characteristics (amplitude and phase) and the normalized cross correlation function. As for the latter, we can see that the correlation peak appears exactly at a lag distance of 5, which is what we would have expected based on how we have generated the second data set. We can also observe that the phase of the FFT points linearly varies with frequency. Thus, we have graphically confirmed the crucial relation between correlator phase ϕ and the delay τ, which is usually expressed as

$$\phi = 2\pi\nu\tau. \tag{10.9}$$

This means that any delay between the two data stream manifests itself as a phase slope in the cross spectrum. In general, we can state that a phase slope in the cross spectrum will lead to a peak in the cross correlation function. Equipped with this knowledge we can now study two other situations, the impact of noise and the effect due to finite bandwidth. If we add a certain amount of noise to the second data set and perform the same analysis as before we obtain slightly different results as depicted in the middle row of Fig. 10.2. As expected, the phase slope in the cross spectrum becomes more noisy, which is not a surprise since we are no longer comparing only time shifted data but need to consider that the datasets are now corrupted by the addition from a certain noise contribution. The other effect

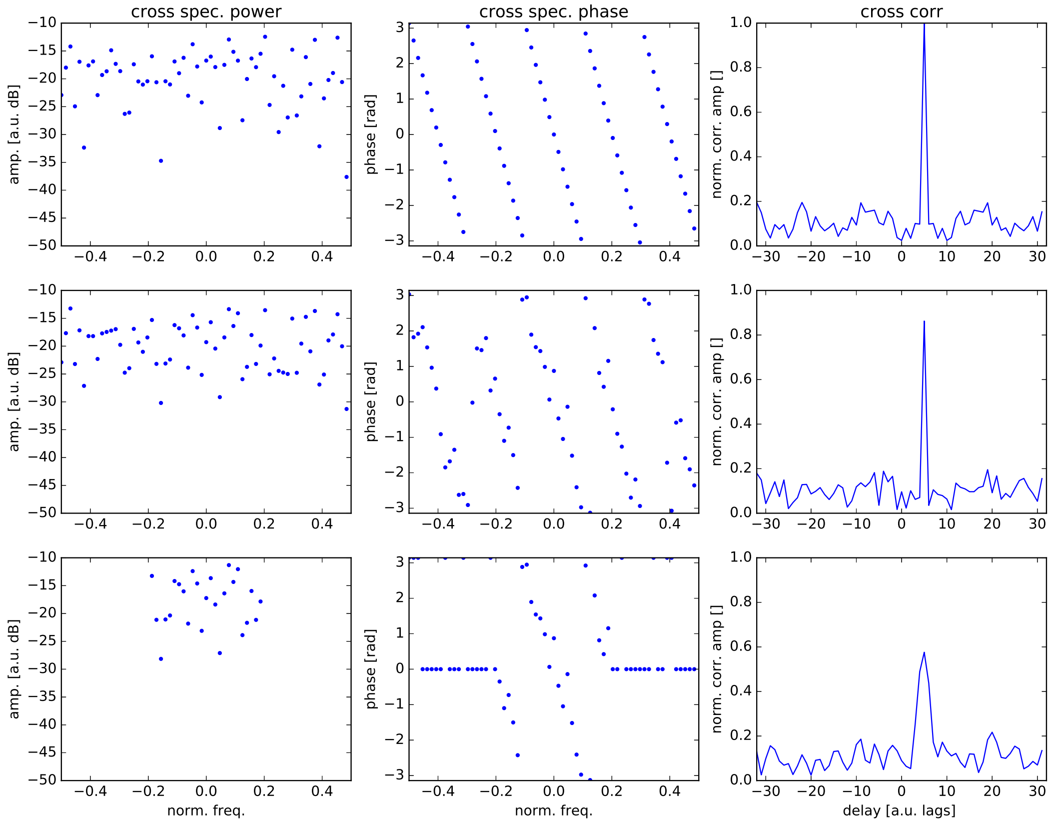


Figure 10.2: Correlation examples based on a synthetic data set. See text for a description of each row's content.

which we can observe is that the correlation amplitude is no longer one as in the case of identical samples, but has decreased significantly. If we define the ratio between the correlation peak and the mean of the correlation function without that peak as the signal-to-noise ratio (SNR) we have gained a very intuitive understanding of how noise actually impact the SNR. Since a large SNR relates to a clear peak detection it is obvious that SNR will later (see Sec. 13) be used as a proxy to describe the uncertainty or precision of a delay measurement.

If we are not only considering noise contribution, but take into account that we are bounded to sample only over a very limited bandwidth, we can carry out another simulation (last row in Fig. 10.2) that depicts this effect very illustrative. Reducing the bandwidth can be easily simulated by setting the complex cross spectrum outside a certain pass band to zero and then perform the inverse Fourier transform to access the cross correlation function. When doing so, we can observe that the correlation peak is still around the 5 lags where we would have expected it to be, but we can also observe that the correlation peak is no longer identified by a sharp peak but is now located on the highest point of what can be approximated well be a quadratic function. In addition, the peak height has further decreased and together with our observation that the peak is now surrounded by a broader range of large values it is again very intuitive that we can deduce a relationship to SNR.

We can now directly relate to what we have defined in Eq. 3.3, where T_{sys} is an equivalent measure of the noise contributions and $\Delta\nu$ represents the effective bandwidth of the receiving system. Lower noise (or lower system temperature) leads to a higher correlation peak and thus to a higher SNR. A wider bandwidth, together with a longer observation time, leads to a clearer phase slope in the cross spectrum and thus a higher and narrower peak in the cross correlation function which in the end translates to a higher SNR or a better formal error of the delay observable.

10.4. Fractional sample delay, fringe rotation and quantization noise

The correlator uses a-priori delay and delay models which allow to account for changes in the delay by dynamically adapting the read address of the raw data streams. The delay is tracked by the correlator using steps in sample units, which causes a difference between a priori delay and the discrete delay realized in the correlator. This difference must not exceed one sample unit within an integration/accumulation period, which normally ranges from roughly one to a few seconds. The difference between the applied discrete delay and the actually modeled delay is referred to as fractional sample delay ϵ and is stored together with the cross spectrum for each integration period. Since we are dealing with cross spectra, we can easily correct for such delays by

$$C'_{12}(\nu) = C_{12}(\nu)e^{-i2\pi\nu(\epsilon_1-\epsilon_2)}, \tag{10.10}$$

where ϵ_1 and ϵ_2 are the fractional delays at stations 1 and 2. Again, we see the advantage of using the cross spectrum since the operation expressed in Eq. 10.10 is computationally not very costly and can be implemented very easily. As shown in Fig. 10.3 one can implement the fractional delay correction directly after the FFT transformation of each station's data stream.

When the delay rate has a positive value the telescope is steadily ”moving away” and thus the Doppler shift causes the signal received to be at lower frequency. Following that same logic, we can state that negative delay rates correspond to Doppler shifts towards higher frequencies. This implies that the signals from two stations cannot be correlated unless the frequency of the signals is referring to the same reference frequency. As depicted in Fig. 10.3, this can be achieved by frequency conversion of the raw sampling data and is easily implemented by multiplication of the data stream with a periodic signal that has a frequency which corresponds to the natural fringe frequency (Sec. 2.5) and which can be considered as the Doppler shift between the two signals. This process is called fringe stopping, and is accomplished by multiplying cosine and sine functions with respect to the time-series data of the stations, i.e., before the data are either correlated or Fourier transformed. Since geodetic VLBI deals with observation frequencies of 8 GHz and above, the delay acceleration $\ddot{\tau}_g$, respectively the Doppler rate, needs to be taken into account as well when applying this correction. Fringe stopping is usually performed at frequencies

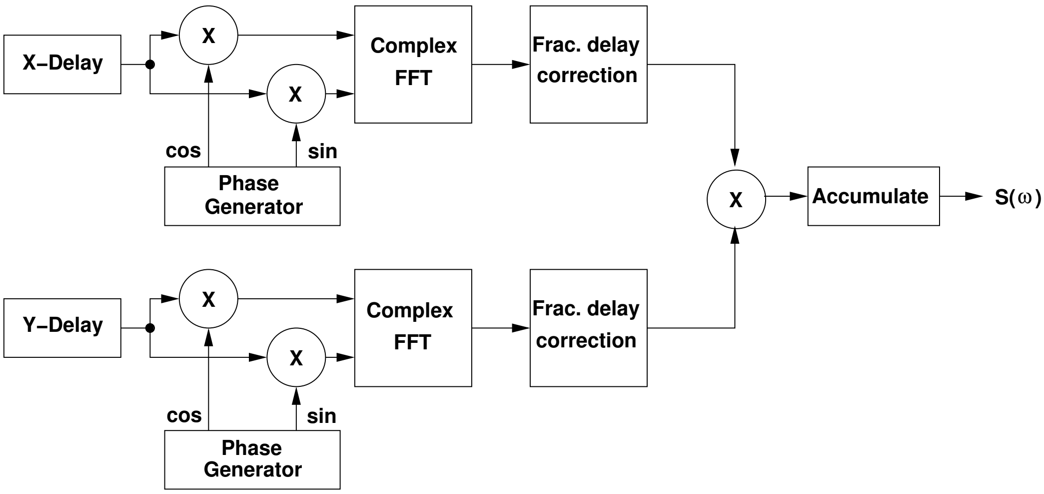


Figure 10.3: Scheme of station-based FX complex correlator according to Whitney (2000) (modified).

calculated at the baseband frequency ν_0 or at the center of the band $\nu_0 + \nu_B/2$ whereas the baseband is defined as the frequency of the radio frequency signal which is finally converted to the zero frequency. If the a priori model is sufficiently accurate, the resulting residual fringe rate will be at the level of a few mHz.

Beside the compensation for finite delay representation and fringe rotation one needs to consider that the recording systems are restricted by their data rate, measured in bits per second (bps). During analog to digital conversion the signal is represented by samples having M quantization levels. Thus, it is obvious that the example discussed before and depicted in Fig. 10.2 would give different correlation amplitudes if we deal with quantized data. However, if the sampling rate and the quantization level is known, which is usually the case for VLBI we can correct the obtained correlation amplitudes for these discretization effect and obtain an unbiased correlation amplitude and thus an SNR value that refers to the original analog signal. According to Van Vleck and Middleton (1966) one can for example correct the correlation amplitude ρ_c obtained from data sets which were 1-bit quantized by

$$\rho = \sin\left(\frac{\pi}{2}\rho_c\right). \tag{10.11}$$

Corrections for other sampling rates or quantization levels can be found e.g. in Thompson et al. (2017).

10.5. Practical considerations of operational correlations

10.5.1. Time settings

To understand the background of the need for accurate time information in the correlation process, we have to look at two things, coherence in correlation and the "UT1-UTC effect". The best results of correlation, i.e., optimum achievable SNR, is reached if maximum coherence is produced. For this, the geometric (telescope coordinates, source positions, EOP) and technical situation (refraction effects and clocks) of the observations has to be reproduced as best as possible. Then the residual fringe/delay rate and residual delay are close to zero and the SNR reaches a local maximum. In contrast to the "UT1-UTC effect", for correlation coherence the information of all clocks in a network only needs to fit in a relative sense. Any change in the absolute value is uncritical as long as it is applied for all telescopes in the same manner in a sense of a Universal Clock Adjustment (UCA)[10].

This is different for the "UT1-UTC effect" which only surfaces numerically at the Level-2 data analysis. The time tags of the observations ($\tau(t)$, $\phi(t)$) implicitly provide the inherent epoch definitions for the parameters to be estimated because the analysis uses the time tags as clock references. For slowly varying parameters such as telescope coordinates, the effect is negligible, but for highly variable parameters, accurate timing information is critical. This is particularly important for the UT1-UTC(t) parameter because t is UTC time and thus the parameter and epoch correlate 1:-1. In other words, any error in t affects the estimated parameter UT1-UTC(t) to the same extent but with the opposite sign as first explained by Clark (1996), later empirically tested in Bonn in 2004 (Himwich et al., 2017), and confirmed by Hobiger et al. (2009).

In the early days of VLBI until the beginning of the 1990ies, it was common habit at the correlators to just assume that one of the telescopes possessed a Hydrogen Maser clock with an epoch setting close enough to UTC. In 1993, the Kokee Park 20 m telescope on Hawaii started operational observations. For convenience and also somehow of personal preference, the KOKEE telescope, or rather the VLBA formatter (Ed Himwich, priv. comm.), was used as the de facto reference for the correlator timing control. The clock section of the correlator control files then listed KOKEE with 0 and the other telescopes relative to this. However, this only applied to those sessions where KOKEE participated. Other sessions were still kind of detached from this convention and the reference times for UT1-UTC determinations may have been inconsistent around that time.

A fundamental change occurred with the deployment of more and more GPS satellites at the beginning of the 1990ies and the development of a low-cost GPS time transfer receiver by Tom Clark, GSFC (Clark, 1996). The Totally Accurate Clock (TAC) was a convenient and unexpensive device for the transfer of GPS time, which by concept had a known and publically available bias to UTC. This equipment was quickly installed at many radio telescopes. Then, the "fmout-gps" readings, linking the formatter/sampler time to GPS time

[10] https://ivscc.gsfc.nasa.gov/publications/newsletter/issue53.pdf, p. 7, Why is my Offset so Peculiar?

(Sec. 5.2.1), could easily be performed with this device and its successors. The readings, in the range of several microseconds, are reported in the log files, which provide the correlators with the necessary information.

Unfortunately, the biases of the station clocks relative to UTC are only part of the story for the epoch definition and subsequently for a reliable UT1-UTC determination. There still are the system delays which originate from inherent delays in the electronics and which cause additional timing biases. Their magnitude is in the range of hundreds of nanoseconds but can easily reach single-digit microseconds although there is a handful of telescopes with biases much later than that. In the mid 1990ies, Kerry Kingham, USNO, and Ed Himwich, NASA, started the pioneering work of monitoring the relative biases, maintaining a list of available numbers, and invented the name "peculiar offset" (PO) for such a value. POs have never been measured directly but were only inferred indirectly because there is a lack of absolute information (Himwich et al., 2017). For this reason, the origin of this list was assumed to be the PO of the KOKEE 20 m telescope with a value of zero. This number applied to the signal chain of the antenna from the beginning of its operation to the middle of 2004 when the VLBA formatter was replaced by a Mark IV formatter (Himwich et al., 2017). The new PO was about (+)400 ns higher than the previous one. For the evolution of the PO following this change, Ed Himwich wrote: "In 2011, the PO was moved an additional +600 ns for the Mark 5B. That give a total of +1000 ns relative to the VLBA formatter value. The application of these changes at the correlators were not perfectly in sync with the formatter actually being used. Then somewhere between 2011 and 2017, the value used at the correlators for Kokee “slipped” to a total value of about +500 ns (that is, -500 ns from where it should be)." The epoch and reason for this change remains in the dark. A graphical representation of the historical developments of the PO of the KOKEE 20 m telescope is available in Himwich et al. (2017).

Monitoring of the POs is carried out regularly and all correlators had agreed to apply this list consistently. The correlator control time section then states the sum of the "fmout-gps" and PO to be applied for initially shifting the epochs of the data. Today, the list of POs is maintained by Sara Hardin, USNO, and made available on a web site[11].

It should be emphasized that the origin of the PO list does not include any system delays which can easily be a couple of microseconds. For this reason, Himwich et al. (2017) assume that there is an overall UT1 bias in the VLBI results "probably in the range of a few μs, positive or negative."

10.5.2. Accumulation period

The software correlator DiFX-2 (Deller et al., 2011), which can be used as a convenient example of a software correlator, an important parameter, normally to be set in a correlation control file, is the length of the integration time. It is often called the accumulation period

[11] https://github.com/USNO-VLBI/config/tree/main/etc

(AP) and determines the duration, in which the VLBI raw delay samples are blocked and for which a single model delay is computed. Although AP lengths are often generalized by 1 to several seconds, the DiFX-2 correlator requires that the length is set to multiples of $0.1\mu s \cdot 2^n$, e.g., $0.1 \cdot 2^{13} = 0.8192s$. However, if the control file specifies an AP length, which does not obey this rule, the software computes and uses the nearest AP length following the rule.

10.6. Extraction of phase calibration information

In the fringe fitting process, we will make use of the information from the artificial calibration signal which was superimposed on the raw VLBI noise (Sec. 7.1.1). This could be extracted at each telescope during or after the observations but the method of choice is to do this during the correlation process. The DiFX-2 software correlator (Deller et al., 2011) performs this phase calibration extraction and writes this information into a separate file for later use. DiFX-2 can be configured to extract any number of tones, unlike many existing earlier hardware implementations which could only provide two tones per sub-band (Deller et al., 2011).

For phasecal extraction, the same data as used for correlation, i.e., streams of the 1-bit or 2-bit sampled raw data are processed in parallel to and entirely independent of correlation. In fact, there are two methods, a standard one and one devised by Walter Brisken.

In the standard method, the first step is a counter-rotation of the sample data for the 10 kHz (legacy) or 400 kHz (VGOS) offsets to place all tones at integer MHz and to produce complex representation. This is done with a pre-computed numerical oscillator of just a few hundred samples. In other words, each component of the vector containing the raw VLBI data in the time domain is multiplied with the corresponding component of a vector v_j consisting of

$$v_j = e^{i\phi_j} \tag{10.12}$$

with

$$\phi_j = -2\pi j \frac{f_{\text{offset}}}{f_s} \tag{10.13}$$

where f_{offset} is the frequency offset of the first phasecal tone and f_s is the sampling frequency/rate, generally two times the channel/sub-band bandwidth (Pogrebenko, 1993a; Pogrebenko, 1993b).

Instead of applying a direct discrete (or fast) Fourier transform (DFT/FFT) of the full accumulation period, which represents averaging over time, the time domain sample data is segmented in N_{M1} bins and the segments are stacked before a DFT/FFT transform is performed at the end. This can be done because the tones repeat. The method is, thus, vastly faster and still provides accurate tone amplitudes and phases (Jan Wagner, priv. comm.).

In this way, the spectral resolution of the Fourier transform is auto-determined because it depends on the sampling rate/frequency and the repetition rate of the phasecal tones in the time domain. The key parameter is the greatest common divisor (gcd) of the sample frequency f_s and the phasecal frequency separation Δf_t. The necessary number N of bins, which determines how many baseband data samples are required before the phasecal time series repeats itself with the same phase, is computed by

$$N_{M1} = \frac{f_s}{\mathrm{gcd}\,(f_s, \Delta f_t)} \tag{10.14}$$

As an example, we may assume an 8 MHz wide channel/sub-band (16 MSamples) and 1 MHz phasecal separation which yields $N = 16$.

In the second method, devised by Walter Brisken, the N-sample segments are much longer, such that they cover one full sine period of the tone offset. Then the gcd of the sample frequency and the frequency offset f_{offset} has to be applied yielding the factor N_{M2} as

$$N_{M2} = \frac{f_s}{\mathrm{gcd}\,(f_s, f_{\mathrm{offset}})} \tag{10.15}$$

For the VGOS case with 32 MHz channel bandwidth (64 MSamples), 5 MHz tone spacing, and $f_{\mathrm{offset}} = 0.4$ MHz, N is 5120.

For both methods, the default in the DiFX correlator software is 1.0 MHz/bin as the minimum spectral resolution of the PCal DFT stage. The length of the DFT gets auto-determined through sampling rate, tone offset from baseband DC, and tone spacing first. For the majority of VLBI setups, this initially determined DFT length is already long enough as-is, and the resolution is usually finer than 50 kHz. However, when not, then the PCal code keeps doubling the DFT length until the bin resolution falls below the configured minimum of 1.0 MHz/bin (Jan Wagner, priv. comm.).

In a more generalized approach, Tasso Tzioumis from some basic considerations for a fine enough resolution after the Fourier transform derives[12]

$$N_{M3} = 200 \cdot 2^n \tag{10.16}$$

for all channel bandwidths 2^n and all tones offset by $k \cdot 10$ kHz for integer k. For a 16 MHz bandwidth and the first tone of a 1 MHz tone frequency being offset by 10 kHz, the segments are N = 3200 samples long which are stored in a vector of this length.

With the vector length determined, subsequent vectors are stacked / added into an N-bin accumulator. After stacking enough for the full integration time, e.g., after 16 such 3200-sample segments of a 1-second accumulation period have been stacked into the accumulator, one does a small N-point complex-to-complex DFT providing a spectrum with 800 complex points. The complex DFT output bins then contain the tone phases and amplitudes.

The phase calibration output file is set up for each scan and each telescope individually

[12] https://www.atnf.csiro.au/people/Tasso.Tzioumis/vlbi/dokuwiki/doku.php/difx/pcalimplicitalgorithm

and contains records for each accumulation period. Each record contains entries for each phase calibration tone consisting of frequency and real and imaginary component. For example, in legacy S/X mode with 16 MHz wide sub-bands/channels, a record may contain 16 tones of 1 MHz separation for each of the 16 S/X frequency sub-bands/channels resulting in 256×3 numbers per record (N.B.: X band also has two lower sideband sub-bands). These numbers are applied to the fringe visibilities at the time of fringe fitting. There are multiple options for this, which are described in Sec. 11.4.

It should be mentioned here that extraction of the phase calibration signal can and should also be done at each radio telescope for diagnostic purposes. Here, the extraction is performed immediately before the digitized VLBI data is stored, i.e., in the digital backend (RDBE or DBBC). This monitoring capability helps to guarantee that the receiver signal is indeed fed through to the end of the signal chain in all sub-bands and no faulty cable connection persists.

11. Fringe fitting and extraction of observables

The next step in the process is the extraction of the phase, delay, and delay rate observables from the correlation output. Correlation is performed so that cross spectra are stored after a certain integration period, often also called accumulation period (AP), which leaves us with a time series of spectra. Knowing that the inverse Fourier transform would give us the correlation peak at each of these integrations, we can now also try to evaluate how the position of this peak varies over time, which in the end will give us the delay rate value that corresponds to the duration of that particular scan. Thus, the process called fringe fitting will use a set of consecutive cross spectra and try to find the most likely values of τ and $\dot{\tau}$ that maximize the correlation function, both in time and in frequency. Plotting all possible combinations of delay and delay rate gives us the so called 2D delay resolution function (see Fig. 11.1) which allows us to identify the most likely values for these two parameters.

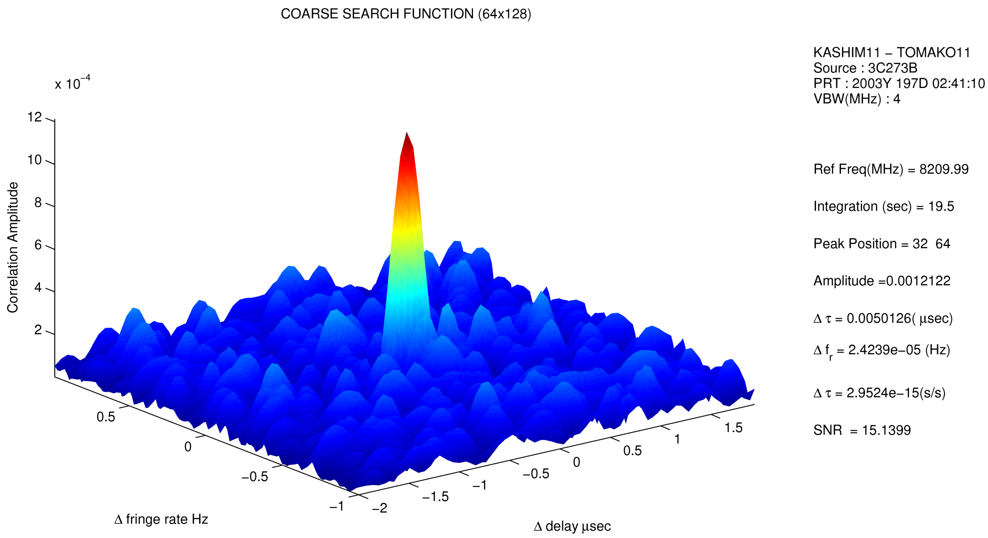


Figure 11.1: Single sub-band coarse delay resolution function (Hobiger and Kondo, 2005)).

For the delay direction we already know that we do not need to manually try all delay values but simply make use of the Fourier transform to obtain the one dimensional delay resolution function. As it turns out, we can also make use of the Fourier transform in the direction of consecutive cross spectra and obtain a peak where we have the most likely value for the delay rate. Thus, the 2D delay resolution function can easily be obtained by 2D FFTs performed on a stack of coherently integrated cross spectra. Once we identified the peak we can do a fine search around that location by a quadratic interpolation and obtain values of τ and $\dot{\tau}$ with a resolution that is better than the corresponding FFT resolution.

As already explained in Sec. 3, the sharpness of the peak depends on the width of the bandpass observed in that the peak is the sharper, the broader the bandpass. However, extremely widening the bandpass also requires more costly data storage and transport.

The path to follow to practically circumvent these limitations is bandwidth synthesis (Sec. 11.5).

11.1. Single band delay

The above considerations all apply to a single solid spectrum of bandwidth B. As we have seen already, modern recording systems always work with multiple, often disjunct bandpasses for the application of bandwidth synthesis (Sec. 11.5). For this reason, we can produce multiple delays of the kind described above. These are commonly called "single band delays" (SBD) but more correctly they should be called "single sub-band" or "single channel" delays.

The practical use of SBDs commonly is, however, not a delay of a single sub-band alone but is the average of the SBDs of all sub-bands/channels in a band. Actually, the complex visibilities or phasors of all sub-bands belonging to a band are superimposed on each other (stacked and averaged) to find the maximum. This process can be considered as incoherent averaging. Consequently, SBD is essentially the delay for the maximum cross-correlation coefficient when all sub-band data of a baseline are correlated.

The average SBD is reported in *fourfit* fringe plots (Appendix F) and is transferred to the databases for Level-2 data analysis. In addition, we can also look up the SBDs of all sub-bands individually and the mean value. These are encrypted in the "Sbd box" (Appendix F).

11.2. Bandwidth synthesis

In order to improve the delay determination under economic restrictions, Rogers (1970) ”solved the problem in a rather ingenious way” - as Whitney (1974) wrote - by developing the bandwidth synthesis method. Similar to antenna aperture synthesis, where a number of small telescopes synthesizes the aperture of a much larger telescope (e.g. VLA or Westerbork Synthesis Radio Telescope), a wide frequency band is sampled using a number of narrow bands, in the case of Mark III/IV/V of 2 to 16 MHz bandwidth. To guarantee coherence, these sub-bands need to be phase-calibrated (Sec. 11.4). Optimization of the sub-band selection in the frequency domain is achieved by placing two sub-bands at the extreme frequencies defining the total spanned bandwidth. In the legacy S/X systems, the separation is 720 MHz at X band and 140 MHz at S band (Fig. 11.2). The target for the VGOS systems is a separation in each band of 1 GHz and 12 GHz between the lowest and highest band. Current VGOS bands are tabulated in Tab. 6.1.

In addition to the outer sub-bands, a number of intermediate frequency sub-bands is introduced in between. While the two extreme frequency sub-bands help to increase the effective bandwidth $\Delta\nu_{RMS}$ according to Eq. 3.12, the interposed sub-bands support this as well but also increase the total bandwidth $\Delta\nu$ in Eq. 3.3. So, using only these narrow sub-bands, the whole bandwidth (between the first and the last frequency) is synthesised

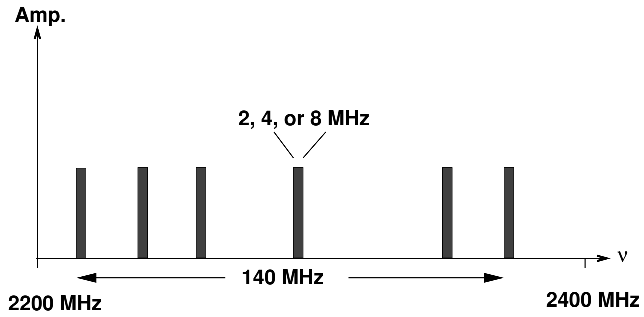


Figure 11.2: Example of frequency sub-band allocation at S band with idealized bandpasses.

and the accuracy of the delay determination is improved. The scheme of sub-band selection is discussed later.

In order to describe the frequency band - time lag (delay) relationship for the bandwidth synthesis case, Rogers (1970) defined a complex delay function for an ideal situation

$$D(\tau) = \int_0^\infty S_{xy}(\omega) e^{-i\omega\tau} e^{-i\omega^0\tau} \frac{d\omega}{2\pi} \tag{11.1}$$

with $S_{xy}(\omega)$ being the cross-spectral function, ω the radio frequency and ω_0 the local oscillator frequency. The envelope or magnitude of this function is called the (one-dimensional) delay resolution function (DRF). The delay resolution function of any synthesis frequency setup has an envelope which represents the result of the Fourier transform of the channel bandwidth similar to Fig. 3.1 with a main peak at 0 delay and first nulls at plus and minus 1/B (Fig. 11.3).

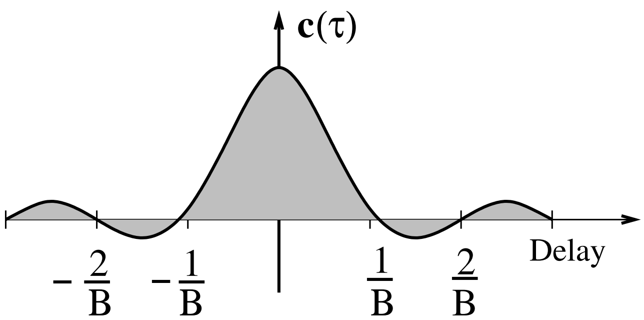


Figure 11.3: Delay resolution function of a monolithic bandpass of width B.

Fig. 11.4 shows the DRF for the case of Mark III with multiple sub-bands of 2 MHz bandwidth. Here, the envelope has a base width of about -500 nsec to +500 nsec (from $(2MHz)^{-1}$). The lower dashed line represents the level of the sidelobes if the total bandwidth would be synthesised perfectly.

Underneath the upper envelope in Fig. 11.4 there are distinguished peaks with sharp edges with one of them being used for the determination of the delay. The sharpness of these peaks depends on the separation of the two outer frequency sub-bands. Their separation and significance result from the selection of the sub-bands in-between the two outer sub-bands. This brings us to the criteria of how and where to place the intermediate sub-bands.

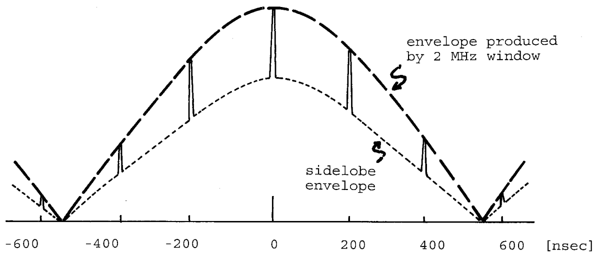


Figure 11.4: Central part of generic delay resolution function with envelope of 2 MHz sub-band bandwidth with main lobe, ambiguous peaks and envelope of sidelobes. Adapted from Rogers (1970).

The starting point of these considerations is the situation where only the two sub-bands at the lower and upper edge are observed. For convenience, we just consider a Mark III S band bandpass spanning 140 MHz. The X band sub-band selection in terms of bandwidth and spacing can easily be derived by the same simple arithmetic. In the above example, there would be peaks underneath the envelope every $(140\,MHz)^{-1}$ or about 7 nsec. The separation of these peaks in the delay domain is called ambiguity spacing. It indicates to what extra delay the identification of the main peak may be shifted if a priori information is not sufficiently good or if electronic channels operate below their specifications. In the bandwidth synthesis case with multiple sub-bands, ambiguity spacings are computed as the inverse of the largest common devisor of all sub-band frequency differences. Since, for a start, we only have the two outer sub-bands, the ambiguity spacing here is just $1/\Delta\nu$.

The 7 ns ambiguity spacing in this example might be too narrow to be solved in the ambiguity resolution step of the level 2 data analysis, and if we use the current X band separation, a spacing of $(720\,MHz)^{-1} = 1.4\,ns$ definitely is. The way to overcome this problem is the introduction of one or more frequency channels in between to suppress the sidelobes.

The basis for the selection of the band separations is the fact that the sharp peaks in Fig. 11.4 are spaced 1/S where S is the greatest common denominator of all n(n-1)/2 frequency differences of n frequency sub-bands used to synthesise the total bandwidth. If, for example, we introduce another sub-band at ν_1 + 10 MHz in between the two sub-bands separated by 140 MHz (i.e., 0, 10, and 140 MHz)[13], the largest common denominator of all frequency differences would be 10 MHz. The resulting separation of the main peaks

[13]This may, for example, represent 2225.99, 2235.99 and 2365.99 MHz at the front end (feed horn and receiver)of the telescope, sometimes also called sky frequency.

then is $(10\,MHz)^{-1}$ or 100 nsec. However, in such a setup, there appears a large number of sidelobes within one ambiguity spacing almost as high as the main peak with a separation of the 7 ns stemming from the 140 MHz spanned bandwidth (Fig. 11.5).

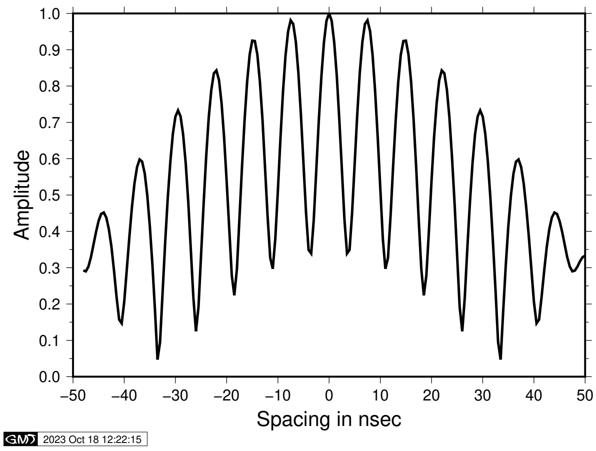


Figure 11.5: Central part of the delay resolution function for a 0, 10, 140 MHz frequency setup. N.B.: This is only the portion between -50 and +50 nsec of Fig. 11.4.

If we consider that the delay resolution function represents the ideal situation with ideal rectangular bandpasses and balanced amplitudes, it is quite understandable that the fringe fitting process, which uses a ”maximum likelyhood method” (Rogers, 1970), may select the wrong peak from several options in Fig. 11.5. The reasons for that may be bandpass irregularities and/or high noise levels. For a better sampling of the whole spectrum, more frequency sub-bands are introduced between the two bandpass edges. They serve the purpose of reducing the level of the sidelobes to a level, which is uncritical for the fringe fitting process, and establishing a practical ambiguity spacing. In addition, more sub-bands provide more total bandwidth and, thus, more bits to be correlated increasing the signal-to-noise-ratio (SNR) which subsequently produces more reliable delay determinations.

The initial selection of the channel separation was done using an algorithm which is based on the Golomb ruler theory (Taylor and Golomb, 1985). A Golomb ruler describes the separation of elements (e.g. frequency bands or antenna separation in an array) on the basis of non-redundant difference sequences, i.e. integer sequences which have the property that the difference of each pair of integers is distinct from the difference of each other pair (Robertson, 1991). However, in practice, technical limitations like radio frequency interferences (RFI), harmonic interferences, and limited receiver bandwidths require compromises in the selection. Therefore, 0, 2, 7, 15, 27, 32, 35, 36 times 20 MHz are used to span the 720 MHz bandwidth of Mark III X band observations and 0, 2, 4, 7, 12, 14 times 10 MHz synthesise 140 MHz of S band. The central parts of the resulting delay resolution functions within the limits of plus/minus one and a half ambiguity spacings are shown in Fig. 11.6 and 11.7. It is immediately obvious that with only 6 channels at S band the

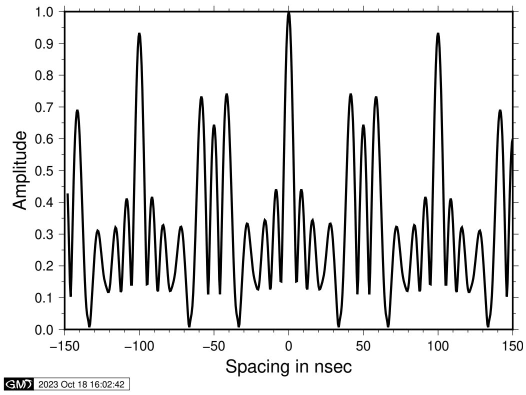


Figure 11.6: Central part of delay resolution function with 6 channels at S band covering 3 times an ambiguity spacing of 100 ns following a sub-band separation of 10 MHz.

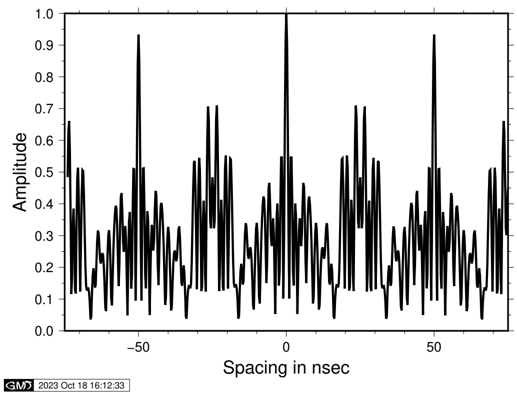


Figure 11.7: Central part of delay resolution function with 8 channels at X band covering 3 times an ambiguity spacing of 50 ns following a sub-band separation of 20 MHz.

suppression of the sidelobes is less successful than with 8 channels at X band.

11.3. Sub-ambiguities and the effects of missing sub-bands

As we have seen in the previous section, sub-bands are introduced in between the two extreme channels for an increase in SNR and for the suppression of sidelobes in the delay resolution function. Since each sub-band scheduled to be observed has its specific ”purpose”, the loss of a sub-band due to radio frequency interference, sampler failure, or other deficits not only reduces the SNR by $1/\sqrt{(n-1)/n}$ but also distorts the delay resolution function. Fig. 11.8 shows the central part of a DRF (just one ambiguity spacing of it) of an otherwise undisturbed S band spectrum with sub-band 3 missing.

It is obvious that the sidelobes within the ambiguity spacing may easily be detected as

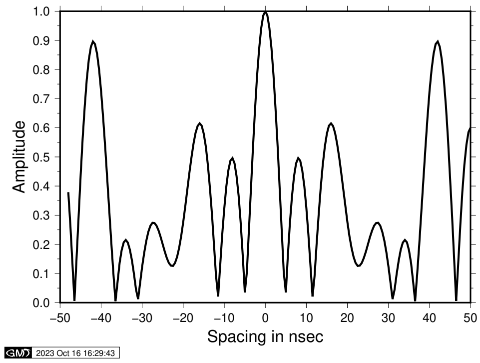


Figure 11.8: Central part of Delay Resolution Function using only S band sub-bands 1, 2, 4, 5 and 6 with 100 ns ambiguity spacing.

the main lobe in case of low SNR or bandpass irregularities. The problem is that there is no closed algorithm to compute, where the sidelobes are located on the delay axis, to apply the necessary corrections.

Another complication arises from the fact, that it may happen that one telescope lacks sub-band x and another telescope lacks sub-band y. Then, on the linking baseline, both sub-bands are missing and the (synthetic) delay resolution function looks as displayed in Fig. 11.9. Although the sidelobes are not that pronounced compared to Fig. 11.8, there may still be a great danger of mis-identification of the main peak in the real situation because of reduced sensitivity.

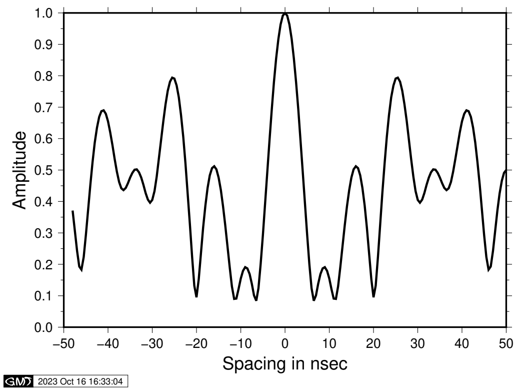


Figure 11.9: Central part of Delay Resolution Function using only S band sub-bands 1, 3, 4, and 5 with 100 ns ambiguity spacing.

The fact that the fringe fitting process did not find the correct main peak cannot be discovered here but only in the level 2 data analysis of all resulting group delays. There, the delays derived in the multi-sub-band fringe fitting process appear as multi-nanosecond outliers. Generally, these observations have to be discarded since their separation from the main lobe may only be computed with low accuracy. However, an iterative method is available, which forces the fringe fitting process to home in on the correct peak. The procedure uses initial group delay residuals and converts them into *fourfit* multiband residual delays (correlator model residuals) and subsequently into search windows of about 10 nsec. With these search windows the fringe fitting process is repeated with a high success rate. However, all scans with missing sub-bands suffer from a reduced SNR which lead to reduced data quality.

Missing channels/sub-bands in raw data of different telescopes may also have an effect on the theoretical ambiguity spacings which may be different for individual baselines in a network. Let us consider the S band channel/sub-band frequencies of Sec. 5.2.2 which were 2225.99, 2245.99, 2265.99, 2295.99, 2345.99, 2365.99 MHz. The largest common divisor of all possible differences is 10 MHz and $1/\delta\nu$ produces an ambiguity spacing of 100 ns. If channel #4 (2295 MHz) is missing in the data of a telescope, the largest common divisor of all possible differences becomes 20 MHz and the ambiguity spacing 50 ns. All baselines with this telescope then have a different ambiguity spacing than the rest.

11.4. Application of phase calibration

For determining the observables group delay, phase delay and phase delay rate in the BWS mode, we are interested in using the information of all sub-bands together and obtain the most likely values which maximize a correlation function that is derived from all these sub-bands. Before we start, however, we have to correct for inter-sub-band phase differences stemming from different electric path lengths between the different sub-bands. For this, we make use of the phase calibration signals which were injected at the first stage of the receiving system to monitor the dispersive delay in the receiving system (Sec. 7.1.1) and which were extracted in the correlator (Sec. 10.6). In other words, we apply a phase shift to each of the sub-band's cross spectra for calibration and then perform a fringe fit over all bands.

The phases (and the amplitudes) of the calibration signals can be extracted by different methods and the results are stored in separate files for application at the time of fringe fitting (Sec. 10.6). The general mode of storing complex phase calibration data for each scan, each telescope, and each accumulation period individually allows for a flexible usage of this information here. However, the many options, we have here today, may blur the view for the general concept of phase calibration.

For this reason, we first depict the logic of phase calibration in simple terms. Let us consider the results of fringe fitting without phase calibration for disjunct BWS frequency sub-bands. The relationship between the phase slope w.r.t. frequency $d\phi/d\omega$ and the delay τ is

$$\tau = \frac{d\phi}{d\omega} = \frac{1}{2\pi}\frac{d\phi}{d\nu}. \tag{11.2}$$

For multiple visibility/fringe phases, the result of the linear regression of the phases across all sub-bands depends on the scatter of these phases and may or may not produce a reasonably good fit (Fig. 11.10). The scatter about the straight line fit depends on the dispersive fringe phase distortions, and in turn, the resulting delay τ has a matching formal error. N.B.: The linear regression is only used as an intuitively easy to understand equivalent to the Fourier transform from the frequency to the delay domain which is used in the routine fringe fitting process. The Fourier transform produces a sharp peak at the correct time delay but this always has an ambiguous repetition.

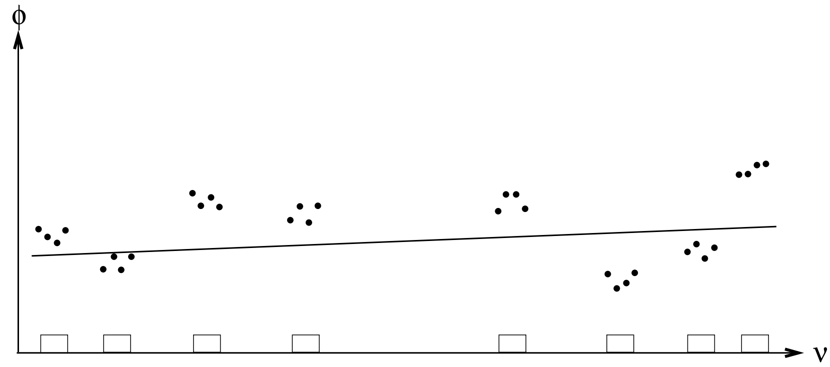


Figure 11.10: Uncalibrated fringe phase versus frequency. The boxes at the bottom depict the limits of the frequency sub-bands/channels while 4 points in each band are examples of the interferometer phases from the cross-power spectrum. The straight line represents the $d\phi/d\omega$ producing the delay τ.

In the same way of representation, the phases of the phase calibration signals are scattered about a straight line (Fig. 11.11). With only one phase calibration tone present in each sub-band, the phasecal phase is assumed constant for this sub-band. Since also for the calibration phases, the relationship $d\phi/d\omega = \tau_{pc}$ applies, we can see that phase calibration provokes an additional delay contribution. It should be emphasized that we always have two appearances of the phasecal phases as in Fig. 11.11, one for the reference and one for the remote telescope, or we can consider the figure as the respective phase differences.

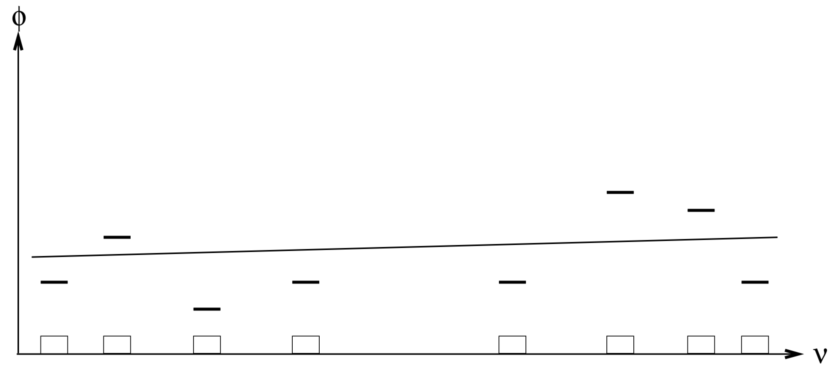


Figure 11.11: Phase cal phases which are constant over each band. N.B.: Since $d\phi/d\omega$ again produces a delay τ_{pc} the straight line again represents a delay contribution.

For quite some time now, more than one phase calibration tone is extracted from each sub-band. In a sort-of phase calibration pre-processing step, multiple tone phases are condensed to a single phasecal phase at the center frequency of the sub-band (Eqs. 11.3 and 11.4). Thus, the same situation applies as we had with the single tone alone earlier on.

Subtracting the phasecal phases of the two telescopes $\phi_{i,cal(1)}, \phi_{i,cal(2)}$ from the fringe phases of our small example of Fig. 11.10 ($\phi_{i,cal} = \phi_{fringe} - (\phi_{i,cal(2)} - \phi_{i,cal(1)})$ then leads to calibrated fringe phases $\phi_{i,cal}$ with a reduced scatter (Fig. 11.12). The fit, resp. the Fourier transform, is then much more stable and the formal error of the delay determination reduces significantly.

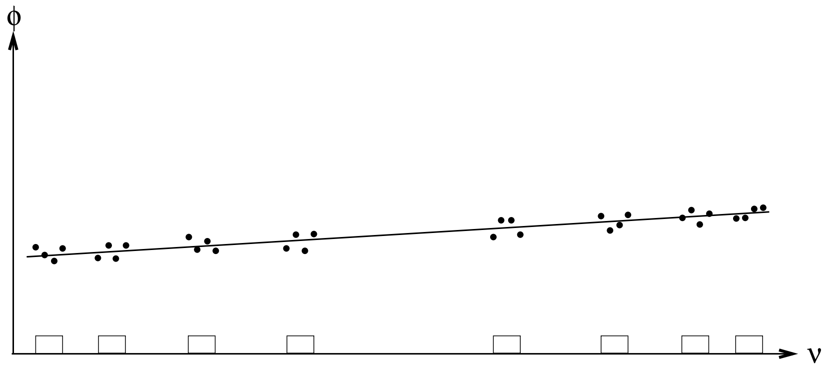


Figure 11.12: Calibrated phase versus frequency. The boxes at the bottom depict the limits of the frequency bands while 4 points in each band are examples of the interferometer phases from the cross-power spectrum. The straight line represents the $d\Phi/d\nu$ producing the delay τ.

In a more detailed description of the application of phase calibration and its pre-processing, we return to the fact that phasecal signal extraction provides arrays of complex phase calibration data. Following HOPS Memo #11[14], we have multiple options to prepare the phasecal input for the calibration at hand with the following keywords in HOPS speech:

normal	obsolete keyword from early Mark III developments with small sub-band bandwidths (up to 4 MHz) using a single tone (2 tones from VLBA correlator) as pre-determined by (hardware) correlator
multitone	the coherent sum of all tones within the channel's bandpass (see below)
ap_by_ap	phasecal phase of a single tone is applied independently for each AP (obsolete through multitone)
manual	one value is manually specified for each channel's phase, often applied if phasecal data is unavailable or corrupt

"multitone" is the most common setup today. As we have seen above, the phase calibration file of a telescope and observation contains complex calibration numbers (phasors) for each accumulation period, each frequency sub-band, and each tone in that sub-band, for

[14]https://www.haystack.mit.edu/wp-content/uploads/2020/07/docs_hops_011_multitone_phasecal.pdf

example, 256 complex numbers for 16 frequency sub-bands and 16 tones each. The aim of phasecal pre-processing is to compute a single calibration phase for the center frequency of each sub-band for each predefined period of time. This is advisable because often the mid-sub-band/channel frequency does not coincide with any phasecal tone frequency.

The general concept is that first a phasecal delay is determined for each sub-band which represents the slope of the phases with respect to the tone frequencies. This could simply be achieved by a linear regression of the phases providing a phase $\phi(\nu_c)$ at the center frequency ν_c and the slope of the phases represented as the phasecal delay τ_{pc} following

$$\phi_i = \phi(\nu_c) + 2\pi \cdot \tau_{pc} \cdot (\nu_i - \nu_c) \tag{11.3}$$

where ϕ_i is the phase of tone i and ν_i the respective frequency.

In *fourfit*, the same logic is followed, i.e., first a best-fit phasecal delay is determined for each sub-band. The averaging period for this is set in the *fourfit* fringe fitting control file in terms of accumulation periods N_{ap}, where N_{ap} can either be a single AP or multiples. Here, the nature of the phasecal information as phasors on a regular grid of equally spaced tone frequencies allows to use FFTs to determine the best fit delays in each sub-band for N_{ap} accumulation periods. The best-fit delays τ_k for n = 1 to N_{tones} tones in a sub-band k are then used to counter-rotate the individual phasecal phasors ρ_n of tone frequencies ν_n to the center frequency ν_c. The sum over all counter-rotated phasors is the calibration phasor P_k of sub-band k (Eq. 11.4).

$$P_k = \sum_{n=0}^{N_{tones}-1} \rho_n e^{i2\pi(\nu_n - \nu_c)\tau_k} \tag{11.4}$$

The phase of the complex phasor P_k is used to correct the fringe phase of sub-band k (Ref. footnote above). This is done block-wise, i.e., the same phasecal phases are applied to every block of N_{ap} fringe phases. Using phasors also incorporates the amplitudes of the individual phasecal signals in the fit which would not be the case in Eq. 11.3.

Applying these phase calibrations to all sub-bands should ideally eliminate the dispersive phase variations of the signal chain and straighten out the phase variations of the fringe phases as depicted in Fig. 11.12.

The choice of N_{ap} for the number of accumulation periods in the averaging process is regularly set to 5 in legacy S/X sessions and 1 in VGOS sessions. The usage of only one AP at a time matches the earlier setup of "ap_by_ap" although with multiple tones now while earlier on only one tone had been used. The choice of integration length mostly originates from stability considerations and the scan lengths which are often much longer for less sensitive legacy radio telescopes.

In the *fourfit* fringe plot, the type of phase calibration handling is depicted under the key "Pcal mode:", e.g., as "MULTITONE, MULTITONE PC period (AP's) 5, 5" (Appendix F). In this example, 5 APs are applied for the data of the reference telescope and the second 5 stands for the data of the remote telescope. The choice of the integration time for the two telescopes could be different but are applied differently only for experimental purposes. The length is a compromise between SNR and the ability to recover variability. By setting the respective number in the *fourfit* fringe fitting control file to 9999, the averaging period is applied to the entire scan length resulting in a single set of phasecal phases for this observation.

The phasecal delays τ_k represent the total delay caused by hardware along the signal path from phasecal injection to digitization. Although phasecal phases are only stored from $-\pi$ to $+\pi$, the slope $d\phi/d\omega$ can, for example, exceed $2\pi/16$ MHz for a sub-band bandwidth of 16 MHz depending on the situation in the fit. The resulting phasecal delay may then be 500 ns or more. For diagnosis, the individual phasecal delays by sub-band are reported in fringe plots as "PC R delays (ns)" for both telescopes separately where R stands for RCP (Fig. App. F.7). In the next line of the fringe plot, the resulting phasecal phases are reported. In both cases, delays and phases, the averages over all calibrations are depicted.

The phase calibration option "manual" is applied if phasecal data is unavailable or corrupt. In this case, the fringe phases of a strong calibrator source with a known flat emission spectrum are analyzed. Any phase variations are attributed to the signal chain and residuals of a linear fit are applied as calibrations with opposite signs. The respective numbers need to be entered in the fringe fit control file.

Manual phasecal needs to be applied with care because it may have an effect on Level-2 data analysis (Sec. 13). There are two reasons: The first one is that any non-zero slope in the phase calibration phases adds an extra delay contribution to the observed delay for every observation in a session. Changing the manual phasecal phases within a session inevitably leads to an artificial clock jump which needs to be parameterized in Level-2 data analysis. The second reason is that manual phasecal phases, which should by nature be constant for the whole observing session, cannot represent any time dependent system delays which will then end up in the clock parameters estimated in Level-2 data analysis.

At the end, it should be emphasized that the phasecal phases across all sub-bands may also have a non-zero slope. This may add an extra delay contribution to the observed delay, although its variations may be quite small. The pre-processing of the phasecal information is done for each observation on a baseline but also in time independently. However, through applying the same calibration phases for each telescope in a single scan, observations in the same scan are correlated and triangles of phasecal delays should close to zero.

11.5. Fringe fitting in bandwidth synthesis environment

The task of the fringe fitting process of multiband data is the same as for single sub-band data. We want to obtain the most likely values of delay and delay rate which maximize a correlation function that is derived from all the sub-bands together. As stated already above, the initial step of the fringe fitting process is the application of corrections for inter-sub-band phase differences stemming from different electric path lengths between the different sub-bands (phase calibration). In other words, we apply a phase shift to each of the sub-band's cross spectra and then perform a fringe fit over all bands.

Finding the multiband delay can be done in two different ways. The first one is that in a first step the cross-spectral function (amplitudes and phases w.r.t. frequency sub-band) is derived from individual sub-bands as described for the single-sub-band case. Then, a Fourier transform is carried out to convert the multiband cross-spectral function into a multiband delay function. The search for the delay basically works in the same way as for the individual sub-bands, i.e., first a rough search for the location of the peak and then a suitable interpolation with selected spectral points in the vicinity of the peak (Whitney, 2000).

The second way of finding the final delay is a joint fit of the single- and multiband delay as well as the delay rate in a common process as realized, e.g., in the PIMA software (Petrov et al., 2011). As we have seen before a wide effective bandwidth increases the SNR and improves the delay precision, which is the ultimate goal to achieve mm to cm precise VLBI delays. However, one needs to reflect on the fact that sub-bands are usually not next to each other in the VLBI frequency bands but are separated by at least several MHz. Thus, instead of one large sinc-like correlation peak we will expect the actual correlation maximum to be surrounded by a certain number of slightly lower peaks, to which one refers to as side-lobes (Fig. 11.13). In general, the sub-band setup is carefully chosen so that one can expect minimum side lobe level in the actual data (cf. Sec. 3).

A further complication arises from the fact that the full bandwidth is not covered with one continuous band and the set of sub-bands do not align to each other without gaps in the frequency domain. For these reasons, we now face the challenge that our obtained delay can be biased by an integer multiple of the base ambiguity spacing. We can thus imagine that the graph Fig. 11.13 repeats every 100 ns which is the base ambiguity spacing for the frequency setup of this example.

To address the ambiguity issue in a bit more detail, we can review the tasks of single- and multi-band delay search also from the aspect of finding a phase slope that connects all the correlation phases of the individual sub-bands. In the case of single band delays, we deal with such a fit in a straightforward way (see discussion earlier) as we have phase values at each of the FFT points and thus should be able to detect any phase slope rather easily if the SNR is high enough. However, in the case of multi-band delay search we have to connect the phases over a much wider bandwidth and the fact alone that we have to deal with gaps between the individual bands, where we have no phase information at all, makes it already harder to find the actual phase slope that corresponds to the residual group delay. On top of that, there is a manifold of possibilities for possible delay slope values, due to the nature of phase measurements being limited to lie between $-\pi$ and π. Thus, we can only determine the residual multi-band delay by satisfying phase slope criteria, but need to consider that our obtained delay value might be off by several integer multiples of the base ambiguity

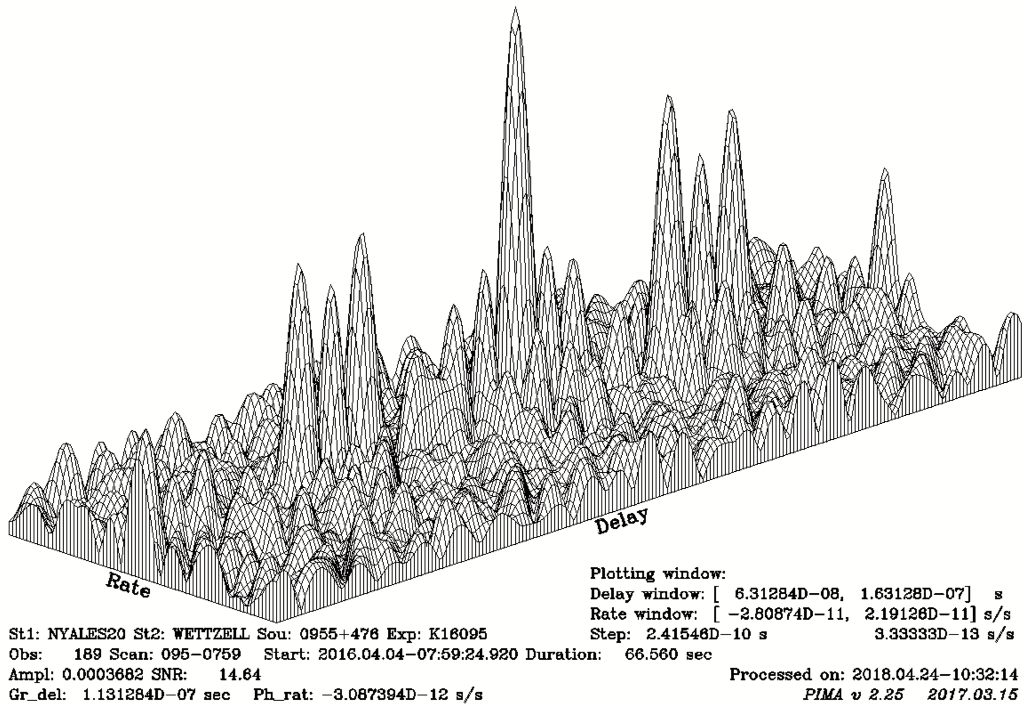


Figure 11.13: Multi-band delay function produced with PIMA (Petrov et al., 2011). S band data with 100 ns ambiguity spacing.

spacing.

As described in more detail in Takahashi et al. (2000) one can find a clear relation between the minimum sub-band spacing and the size of that base ambiguity. In general, one can state that the ambiguity spacing τ_{amb} corresponds to the greatest common divisor of all sub-band spacings $\Delta\nu_{max}$ by

$$\tau_{amb} = \frac{1}{\Delta\nu_{max}}. \tag{11.5}$$

Considering for example the eight USB sub-bands of the standard geodetic X band setup (Sec. 3), we find a $\Delta\nu_{max} = 20$ MHz. Thus we face an ambiguity spacing of 50 ns, which is fortunately rather easily being dealt with given that the a priori information in the data analysis is good enough to provide theoretical delays with an uncertainty that is not exceeding half of the ambiguity spacing, i.e. 25 ns or about 7.5 meter.

11.6. Extraction of total phase and phase delay

So far, we have only looked at the group delay which is found in a 3D search in the single-band delay, delay rate and multi-band delay domain. For the derivation of the phase observable, we have to distinguish between total phase with respect to a reference frequency ν_0 in radians or degrees $\phi_{total}(\nu_0)$ and the phase delay in time units τ_{ph}.

The basis for these observables are the phases of the cross-power spectra $\phi(\nu,t)$ which appear across the visibility frequencies in a single sub-band as well as across the visibility frequencies of all bandwidth synthesis sub-bands. The number of these visibility phases depends on the spectral resolution of the Fourier transform of the VLBI raw data as set for the correlation process. The basic model of the observed phase as a function of frequency for a certain epoch t can be formulated as

$$\phi(\nu,t) = \phi_0(\nu_0) + \frac{\partial\phi}{\partial\nu}\cdot(\nu-\nu_0) + \frac{\partial\phi}{\partial t}\cdot(t-t_0) \tag{11.6}$$

with ϕ = phase in cycles, ϕ_0 = phase at ν_0 in cycles, ν = frequency in Hz, ν_0 = reference frequency in Hz.

For the extraction of the phase observable, all phase calibrated (see Sec. 11.4) visibility phases, first, need to be counter-rotated, one by one, according to Eq. 11.6 yielding the counter-rotated phases $\tilde{\phi}(\omega_0,t)$. For this step, we need the a priori model plus corrections for the delay and rate estimated in fringe fitting process yielding τ_{obs} and $\dot{\tau}_{obs}$. This step is done not only across spectral channels but also across all accumulation periods. The result is that all phases are now more or less constant across time and frequency. In the next step, these can then be averaged over time and frequency independently first to allow for a normalization over time in case the individual BWS sub-bands cover different data spans. In principle, the result is the coherently averaged phase $\phi_{coh}(\nu_0)$:

$$\phi_{coh}(\nu_0) = \frac{1}{m}\sum(\frac{1}{n}\sum\tilde{\phi}(\nu_0,t)) \tag{11.7}$$

with n = number of accumulation period in each sub-band and m = number of sub-bands.

As a side line, it should be mentioned that the fringe fitting software *fourfit* for this purpose maximizes the magnitude of a complex sum of visibility phasors. This in essence produces a weighted (by the magnitude of each phasor) sum that determines a mean phase (Capallo, priv. comm.). For more details see Whitney (2000).

To acquire the phase delay, first the (correlator) residual phase delay $\Delta\tau'_{ph}$ is computed

$$\Delta\tau'_{ph} = \frac{\phi_{coh}(\nu_0)}{2\pi\cdot\nu_0} \tag{11.8}$$

and then added to the a priori delay τ_g

$$\tau_{ph} = \tau_g + \Delta\tau'_{ph} \tag{11.9}$$

For the total phase $\phi_{total}(\nu_0)$, first the a priori phase $\phi_0(\nu_0)$ is computed

$$\phi_0(\nu_0) = \nu_0\cdot\tau_g\cdot 2\pi \tag{11.10}$$

which is then added to the coherently integrated phase $\phi_{coh}(\nu_0)$

$$\phi'_{total}(\nu_0) = \phi_0(\nu_0) + \phi_{coh}(\nu_0). \tag{11.11}$$

$\phi'_{total}(\nu_0)$ is an unambiguous total phase with many phase turns. Before *fourfit* displays and exports the total phase, it applies the modulo 2π function

$$\phi_{total}(\nu_0) = mod(\phi'_{total}(\nu_0), 2\pi) \tag{11.12}$$

which makes $\phi_{total}(\nu_0)$ ambiguous.

The frame of the a priori delay τ_g can either be chosen to be the geocenter delay or the station-based delay (see Sec. 2.4). Depending on this, τ_{ph} from Eq. 11.9 and $\phi_{total}(\nu_0)$ from Eq. 11.11 are either geocenter or station-based phase delays and total phases, respectively (Corey, priv. comm.).

For completeness, it should be added that for VGOS broadband observations, the dispersive effect of the ionosphere has to be added to Eq. 11.6. The additional term is $[-1.3445/\nu \cdot \Delta TEC]$ where ΔTEC is the differential Total Electron Content in TEC units (1 TECU $\equiv 1 \times 10^{16}$ electrons). In *fourfit*, the actual peak of the magnitude of the coherent sum is found by parabolic interpolation of magnitudes at closely-spaced grid point values for the independent variables group delay, single band delay, group delay rate, and dTEC (Capallo, priv. comm.).

12. Special delay considerations

As has been mentioned already briefly at various places of the document, multiple different delays exist. These are summarized here together with a few more hints on some distinctions and special relationships such as multiband - singleband delay differences and triangle closures. In general, for clarity the delay observables would require further specifications or descriptive attributes but often the expression delay is only generalized.

12.1. Types of delays

As already explained in Sec. 2.2 of the basic VLBI theory, the initial observed VLBI quantity is a phase delay because it is conceptually based on a single, monochromatic wave. Changes in this phase over time lead to the phase delay rate (Sec. 2.3). Since in geodetic and astrometric applications, we always observe a band of frequencies with a certain bandwidth, narrow or wide, the delay refers to a group of frequencies, thus the expression group delay. All our time delay observables are group delays.

We have also seen in Sec. 2.4 that the choice of reference epoch is important. We distinguish baseline delays with the reference epoch being referred to one telescope, sometimes also called station delay, and geocenter delay with the reference epoch being the one when the wavefront passes the geocenter. Again, there may be phase and group delays.

Originating from the way how fringe fitting is performed, we also distinguish between singleband delay and multiband delay. As we have seen in Sec. 11, multiple slices of the electromagnetic spectrum are observed and recorded. At the time of the Mark III era, each of these sub-bands/slices was processed by a separate chain or channel of electronics. To describe how the delay was derived, it was either called singleband delay or multiband delay. The singleband delay is computed from an incoherent sum of all fringe phasors/fringe visibilities over all sub-bands/channels. So, it is the average of all individual sub-band delays. The multiband delay is determined from multiple sub-band channels in a bandwidth synthesis approach (Sec. 11.2). This means that phase calibration is applied first and fringe fitting makes use of the full spanned bandwidth in a coherent way.

This distinction is also of importance in the new VGOS era, although in a slightly different meaning. Again, every band (A, B, C, or D) has its own single band delay. However, in contrast to S/X legacy mode, phase calibration has been applied beforehand and the VGOS singleband delay is derived as a bandwidth synthesis delay. It uses the data of all 32 MHz wide sub-bands sampling the total bandwidth of 512/1024 MHz for each of the VGOS bands A, B, C, or D separately. VGOS multiband delays are a result of a coherent integration across all sub-band samples of all bands and of a 3D maximum search (Sec. 11) after phase calibration has been performed (Sec. 11.4).

12.2. Relationship of singleband and multiband delays in legacy systems

Although it may be intuitive wisdom that the singleband delay (SBD) and the multiband delay (MBD) of the same observation in legacy S/X systems should agree within a few nanoseconds (the average precision of SBD), this, unfortunately, is not correct. The main driver for the differences is phase calibration. Let's have a look at the signal flow in a legacy hardware setup and here especially at the different delay contributions. For this description, I have to give special credit to Brian Corey who explained to me most of what I am writing here.

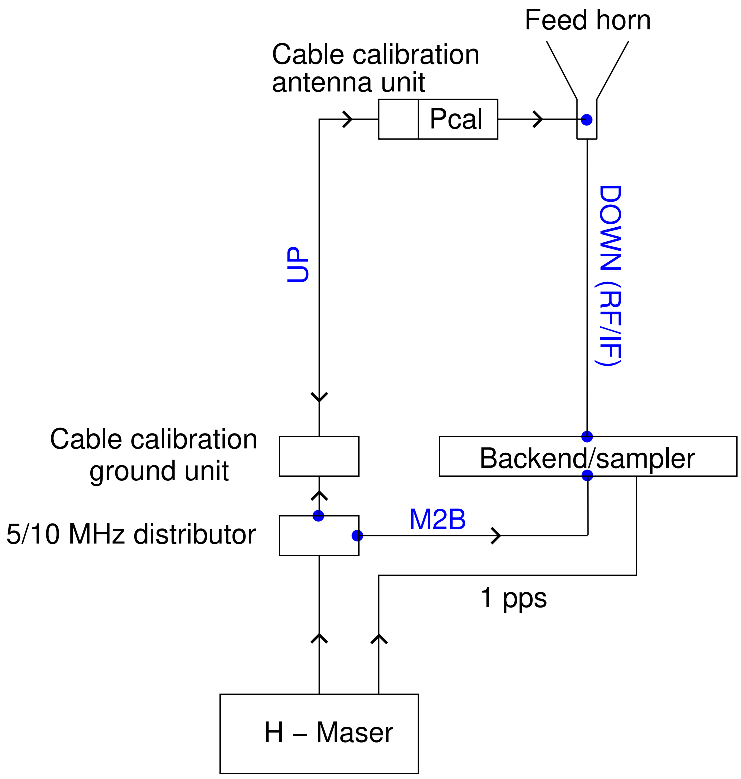


Figure 12.1: Individual delay contributions in a legacy S/X system.

A simplified description of the delays at a telescope has four main contributions

MAIN delay	geometry + atmosphere + clock, exclusive of delay from receiver feed to backend
UP delay	delay in phasecal reference cable (usually 5 MHz) from maser to receiver
DOWN delay	RF/IF delay from receiver feed to backend
M2B delay	5 MHz delay from maser to backend

The MAIN delay can be considered as a delay with respect to the geocenter. This facilitates the logic when we form a baseline with two telescopes. Following Fig. 12.1, the total signal delay at the backend of telescope (i) equals MAIN_i + DOWN_i. At the processing stage, the fringe fitting programs such as *fourfit* correct the sub-band data for the phase-cal phases, which produces an extra phasecal delay. This delay originates from the signal delays in UP, DOWN, and M2B through

$$\text{Phasecal}_{(i)} = UP_{(i)} + DOWN_{(i)} - M2B_{(i)} \tag{12.1}$$

and has to be subtracted, so that the contribution of telescope (i) to $\text{MBD}_(i)$ is

$$\begin{aligned} MBD_{(i)} &= (MAIN_{(i)} + DOWN_{(i)}) - \text{Phasecal}_{(i)} \\ &= (MAIN_{(i)} + DOWN_{(i)}) - (UP_{(i)} + DOWN_{(i)} - M2B_{(i)}) \\ &= MAIN_{(i)} - UP_{(i)} + M2B_{(i)} \end{aligned} \tag{12.2}$$

In general, SBD is not corrected for the phasecal delay. Consequently, we only have

$$SBD_{(i)} = MAIN_{(i)} + DOWN_{(i)} \tag{12.3}$$

and the difference is

$$MBD_{(i)} - SBD_{(i)} = M2B_{(i)} - UP_{(i)} - DOWN_{(i)} \tag{12.4}$$

UP and DOWN are basically the cable/fiber delays between control room and receiver and can easily be hundreds of nanoseconds. This is reflected also in the constant part of phasecal (Sec. 11.4) but not so much in the variable part which normally is only in the picosecond regime. Hence the total MBD - SBD difference (originating from the difference of two telescopes) can take on large values. However, the fringe fitting process limits this to the range of plus or minus half a multi-sub-band ambiguity spacing.

In addition to this information, there are a few other facts. As can be seen, the delays originate from individual telescopes. For this reason, the MBD - SBD differences generally close in triangles but may mis-close by one or more ambiguity spacings. MBD - SBD differences can occasionally be close to zero if the cable delays of the two telescopes agree. S and X band values of MBD - SBD differences differ because S and X electronics and cables differ, and hence S and X values of DOWN differ.

It may happen that there are significant jumps in MBD - SBD differences. If the jump is the same for S and X, such a jump would be caused by a change in M2B and UP. The DOWN links are different for the two frequencies. A jump should not, however, be caused by a real Maser clock jump, as that would affect SBD and MBD equally.

12.3. Relationship of singleband and multiband delays in VGOS systems

What was written above for the legacy systems basically applies also to VGOS processing. However, in contrast, the real SBD (for band A,B, C, or D) is corrected for phasecal delay, and MBD - SBD differences are much smaller. However, application of phasecal is not performed in the conventional, BWS sense as for MBD. In fact, when multitone phasecal is specified for a station in the *fourfit* control file, a phasecal delay is calculated from the phases of the tones within and for each channel. Then, the station differences in these phasecal delays are used to correct the channel SBDs. The intention is to remove instrumental effects within the channels (Brian Corey, pers. comm.).

In practice, an additional requirement for the application of multitone delays is that sampler pools be defined which comprise channels that share a common sampler. Then, the averaged tone-derived differential delays are applied to all channels sharing a sampler. For example, the band A channels, which share a common sampler, have their spectral phases shifted by the mean phasecal delay over the eight band A channels. The same applies for the other three bands. The SBD for all 32 channels is then estimated from incoherent average over the 32 channels. "It is the inclusion of 'samplers' in a VGOS control file (along with multitone pcal) that leads to fourfit correcting SBD for channel pcal delays. S/X control files do not use 'samplers', so no SBD correction is done, even when multitone is used." (Brian Corey, pers. comm.)

As a consequence of this, and of course of the fourfold number of channels, the VGOS SBD is a factor of 10 more accurate than the SBD of legacy setups (1σ 250 ps vs. 2.5 ns). For this reason, the ambiguous VGOS multiband delays can easily be successfully controlled to stay close to the SBD by the *fourfit* control command 'mbd_anchor sbd' (see Sec. 12.4).

12.4. Group delay ambiguities

As has been explained in Sec. 11.2 on bandwidth synthesis, group delays derived with multi-sub-band/-channel processing face the drawback of being ambiguous by tens of nanoseconds. The actual number, in S/X/K band observations mostly modulo 10 ns, depends on the selection of the frequency setup. VGOS observations currently have a fixed ambiguity spacing of 31.250 ns (1/32 MHz). In S/X band observations, the appearance of ambiguities in the observed delays transferred to the Level-2 data analysis leads to some extra work for the analysts. Although generally rather trivial because the ambiguity contributions are large enough to be found, clock parameters need to be estimated first. This is done either with delay rate observations, a long forgotten trick of the last century, or with singleband delays. The multiband delay residuals computed with these clock parameters show distinct ambiguity levels which can be used to apply ambiguity corrections to the multiband delay observations. The complication arises from the fact that the correct corrections must not compromise the group delay closure condition (Sec. 12.5). For this reason, sometimes the need arises that data needs to be checked manually.

As it happens, the propagation of ambiguities into the total observed group delays, not its original nature, originates from the handling of the residual delays determined in the fringe fitting process and the addition of the theoretical delays used in the fringe fitting program. Since the majority of the VLBI sessions are processed routinely with the HOPS *fourfit* program, we find a clue in the *fourfit* control file[15]. Here, the keyword "mbd_anchor" defines what *fourfit* uses as the reference for the decision to what reference information the final observed multiband group delay is shifted. The option "model" is the default and the observation is shifted to the correlator model delay to within half an ambiguity spacing. The second option is "sbd". Here the observed delay is shifted to match the singleband delay to within half an ambiguity spacing.

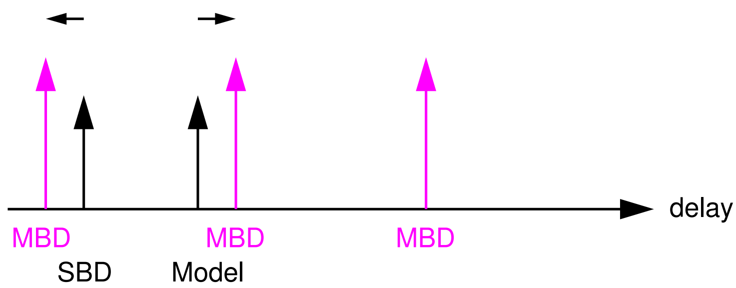


Figure 12.2: Relationship of SBD, model delay, and MBD.

For the processing of VGOS data, the *fourfit* control files always contains the line "mbd_anchor sbd". As a result, when the residual delay drifts over a session due to an inaccurate clock model, the total MBD should track that drift without ambiguity jumps because the (ambiguity-free) SBD tracks the drift.

For S/X, speculation is that the default is used because there are no *fourfit* control files with the "mbd_anchor" action keyword. This means that the total MBD is set to the value closest to the model delay and ambiguity jumps may then occur if the model delay does not match reality well enough. In this respect, MBD ambiguities of S/X observations are pure *fourfit* artefacts which may not appear with other fringe fitting programs such as PIMA (Petrov et al., 2011).

12.5. Triangle delay closures

For various purposes, the forming of triangles of group or phase delay observables helps to monitor certain characteristics of VLBI observing sessions and possibly identify errors. In general, it can be stated that triangle delays should close to zero except for source structure effects which are source and baseline dependent. In addition, systematics through the loss of channels may also lead to mis-closures. Since these are purely baseline dependent, they

[15] https://www.haystack.mit.edu/wp-content/uploads/2020/07/docs_hops_009_fourfit_users_manual.pdf

may be detectable at the level 2 data analysis stage through so-called baseline-dependent clock offsets (BCOs). All other effects are telescope dependent and cancel if the direction of forming the triangle and adding the delays is performed correctly. When assessing the statistics of triangle closures, it is important that for all triangles formed the same direction of rotation is maintained at all times, i.e., either always clock-wise or always anti-clock-wise. Otherwise the distribution function of the closure delays is corrupted.

For the following considerations, we have to distinguish between the geocenter delays and baseline delays. This differentiation primarily refers to the reference epoch of the observables.

12.5.1. Geocenter delay

The invention of the geocenter delay originates from the necessity of computing a priori delays for the correlation process. It is quite obvious that this process requires only $n+1$ delay computations, i.e., for n telescopes plus for the geocenter, and the necessary differencing for the baselines (Eq. 2.8). This is much less computational effort than computing an a priori delay for each baseline, which increases the computational load approximately with the square of the number of telescopes ($n \cdot (n-1)/2$). In addition, it should be emphasized that the delays are automatically referred to the same epoch (t_{GC}) while in the case of the baseline delay (see Sec. 2.4) a different reference epoch might exist for each baseline. The identical reference epoch brings with it the advantage that the closures can be formed directly:

$$\tau^{gc}(close) = \tau_{(1)(2)}(t_{GC}) + \tau_{(2)(3)}(t_{GC}) + \tau_{(3)(1)}(t_{GC}) \tag{12.5}$$

As long as the reference epoch is kept identical, the relationship

$$\tau_{(1)(3)}(t_{GC}) = -\tau_{(3)(1)}(t_{GC}) \tag{12.6}$$

remains intact and closures can be computed simply from the delay values reported in any order of the telescopes just observing the indexing and the direction of the initial baseline vectors. $\tau^{gc}(close)$ should be zero but mostly is not because of the baseline dependent source structure effects or unsymmetric channel losses.

12.5.2. Baseline delay

Correctly, the baseline delay should be called station-based baseline delay because the epoch of the delay is referred to the time when the wave front passes one of the telescopes. In the example in Fig. 12.5, the wave front arrives at telescope (1) before reaching telescope (2) at $t_{(2)}$ forming the delay $\tau_{(1)(2)}(t_{(1)})$. The same applies to the other two delays in the triangle.

The baseline delay is the delay which is the input for any level-two data analysis for

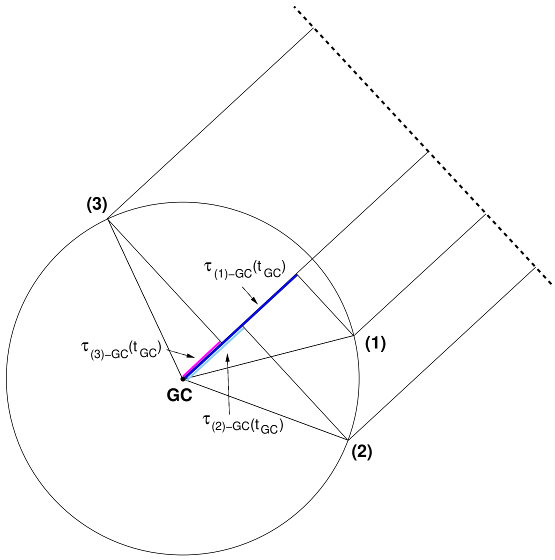


Figure 12.3: Individual geocenter delays $\tau_{(i)-GC}$ at time of arrival of the wave front at the geocenter (t_{GC}).

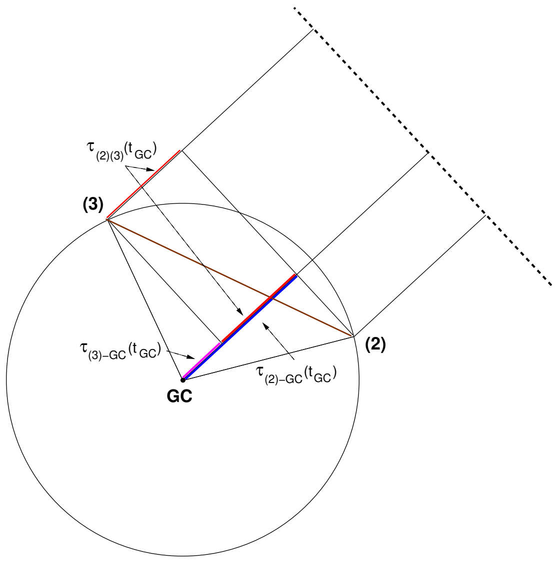


Figure 12.4: Forming of a baseline geocenter delay (red line) as the difference of the two telescope geocenter delays (blue and magenta lines).

estimating geodetic parameters. However, as stated above the primary output of modern software correlators, such as DiFX, are geocenter delays. For this reason, in an intermediate step, the geocenter delays $\tau^{gc}_{(1)-(2)}$ are converted to baseline delays $\tau^{bl}_{(1)-(2)}$ (Corey, 2000) using the formula

$$\tau^{bl}_{(1)(2)} = \tau^{gc}_{(1)(2)} - \dot{\tau}^{gc}_{(1)(2)} \cdot \tau'_{(1)}(1 - \dot{\tau}'_{(1)}(t_{gc})) + \frac{1}{2}\ddot{\tau}^{gc}_{(1)(2)} \cdot \tau'^{2}_{(1)}(t_{gc}) \tag{12.7}$$

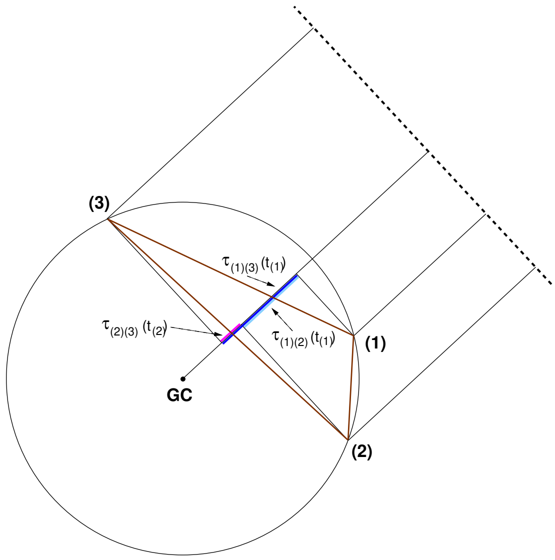


Figure 12.5: Station-based baseline delays referred to the epochs when the wave front arrives at the telescope closest to the radio source.

which follows a Tailor expansion in the derivation. $\tau'_{(1)}$ is the model delay (usually from the CALC module) without the clock contribution for telescope (1) with respect to the geocenter. $\dot{\tau}^{gc}_{(1)(2)}$ and $\ddot{\tau}^{gc}_{(1)(2)}$ are the delay rate and delay acceleration (first and second time derivatives of the delay), respectively.

The computation of the closure delay for station-based delays is slightly more complicated than that for the geocenter delays. The reason is the reference epoch of the individual delays in the triangle. With $t_{(1)}$ and $t_{(2)}$ being the reference epochs when the wave front passes through telescopes (1) and (2), respectively, the closure delay $\tau^{bl}(close)$ is

$$\tau^{bl}(close) = \tau_{(1)(2)}(t_{(1)}) + \tau_{(2)(3)}(t_{(2)}) + \tau_{(3)(1)}(t_{(1)}) \tag{12.8}$$

Note the different epoch for $\tau_{(2)(3)}(t_{(2)})$. In practice, the fringe fitting programs, such as *fourfit* of the Haystack Observatory Postprocessing System (HOPS)[16], compile the reference epochs for all observations of a multi-telescope scan of the same radio source to have the same observation epoch (The time stamp $t_{(1)}$ can then be any other suitable epoch and the use of $t_{(1)}$ does not hurt the following relationships). Anyway, delays with identical epochs in a triangle cannot be summed up as in Eq. 12.8 regardless because this would ignore the fact that the wave front in Fig. 12.5 still travels some finite time past telescope (1) before it arrives at telescope (2). The reference epoch for the delay $\tau_{(2)(3)}$ is thus $t_{(2)} = t_{(1)} + \tau_{(1)(2)}$.

[16]https://www.haystack.mit.edu/haystack-observatory-postprocessing-system-hops/

For the closure delay $\tau^{bl}(close)$ this leads to

$$\tau^{bl}(close) = \tau_{(1)(2)}(t_{(1)}) + \tau_{(2)(3)}((t_{(1)}) + \tau_{(1)(2)}) + \tau_{(3)(1)}(t_{(1)}). \tag{12.9}$$

During the small time lapsing, $\tau_{(1)(2)}$, the Earth continues rotating and telescope (2) moves a finite distance. For this reason, the delay $\tau_{(2)(3)}(t_{(2)})$ will be different from that reported for epoch $t_{(1)}$, again applying a Taylor expansion,

$$\tau_{(2)(3)}(t_{(2)}) = \tau_{(2)(3)}(t_{(1)}) + \dot{\tau}_{(2)(3)}(t_{(1)}) \cdot \tau'_{(1)(2)} + \frac{1}{2}\ddot{\tau}_{(2)(3)}(t_(1)) \cdot \tau'^{2}_{(1)(2)} \tag{12.10}$$

This is almost invers to Eq. 12.7 and again the last term can safely be ignored since it is always smaller than $1 \times 10^{-14}s$. $\tau'_{(1)(2)}$ is the theoretical delay. The closure delay relationship in the station-based delay case (Eq. 12.9) can then be formulated equivalently (ignoring the accelerations), to

$$\tau^{bl}(close) = \tau_{(1)(2)}(t_{(1)}) + \tau_{(2)(3)}(t_{(1)}) + \tau'_{(1)(2)} \cdot \dot{\tau}_{(2)(3)}(t_{(1)}) + \tau_{(3)(1)}(t_{(1)}) \tag{12.11}$$

For closure delay analyses, it has to be emphasized here that the delay $\tau'_{(1)-(2)}$ is the theoretical delay and must not be taken from the listing of the observables because this includes the clock biases which are not applicable. The same of course also applies to any ambiguity offsets remaining in the original observations.

13. Level-2 data analysis

Geodetic and astrometric data analysis of the observations has a lot to do with the terrestrial and celestial reference frames (Sec. 13.1.1 and 13.1.2) as well as with the variable rotation of the Earth parameterized through Earth orientation parameters (EOP). For this reason, we start with the details of reference frames and EOP.

13.1. Reference frames

The concept of reference frames is one of the important advancements of geodesy to describe the kinematics of objects. In space geodesy in particular, the determination of reference frames is one of the principle goals of many endeavors. In early math classes as a first step, we mostly drew a set of coordinate axes with the origin at the intersection of the axes. Then we identified locations of points by measuring distances along the axes starting at that origin to find the respective coordinate components.

In space geodesy, we can neither localize the origin nor figure out where the axes point to a priori. For this reason, we have to approach the problem from the opposite. We assign coordinates to points and infer indirectly where the origin is and where the axes point to. The selection of the points and of the coordinates is done by convention, i.e., by decision as long as the internal configuration, which is the relative relationship between the points, is not deformed (Sec. 13.8). The same also applies to the kinematic movements of these points because we live on an Earth which is changing constantly.

The consequence of this procedure is that a number of points on Earth with their physical markers, which in our case are the reference points of radio telescopes, and their coordinates and all time variable coordinate components define the origin and axes of a terrestrial reference frame (TRF). This means that within its formal errors the origin and axes are fixed through the parameterization chosen for this reference frame, e.g., positions at a certain epoch, linear velocities, post-seismic deformations, and seasonal variations. The more points indirectly fix the frame, the better. Consequently, any telescope can move freely in that frame and the kinematics are described in that reference frame.

The concept of the Global Geodetic Observing System (GGOS) calls for a target accuracy of 1 mm and 0.1 mm/y (Plag and Pearlman, 2009). This should be understood in the way that the origin and the direction of the axes are supposed to be stable with these uncertainties for the next decades to come. The reason is that any phenomena of Global Change, especially global sea level rise, will have to be identified and quantified with respect to the GGOS reference frame or its predecessors until the GGOS requirements are met.

In the sky, the work on the celestial reference frame (CRF) follows the same concept and aims at similar stability criteria. The only difference with respect to the TRF is that the origin of the CRF is defined a priori by the solar system barycenter.

13.1.1. Terrestrial reference frames

At present, the network of radio telescopes for geodetic and astrometric observations consists of approximately 40 active VLBI radio telescopes. The coordinates of these and all radio telescopes active in the past are referred to a conventional Cartesian terrestrial reference frame, although they could also be referred to any other practical terrestrial reference frame. Conventional means that beforehand some geometric guidelines have to be provided by some responsible organisation such as the International Association of Geodesy (IAG) or the International Earth Rotation and Reference Systems Service (IERS). For example, these define, where the origin is supposed to be and where the axes point to, in addition to fixing some definitions and constants (See IAG Resolutions for more details).

More background of conventional terrestrial global reference frames can be understood through looking at the historical developments. So, now we first go back in time.

When the first evidence appeared that the Earth was a sphere and that there was a rotation, which was first seen as the celestial objects rotating around Earth, astronomers developed the concept of a rotation axis, though fixed in space at that time. Together with the rotation axis and the sphere, it was only consequent that the sphere had two poles, an equator and many latitude circles which could actually be measured by star culminations and which helped to interpolate between the poles. While the poles are two points, which are defined geometrically and through physics, the second dimension, the longitudes, caused more complications. Where is zero longitude? For convenience, every state just defined its own zero longitude for navigation purposes. Only in the year 1884 with the International Meridian Conference at Washington D.C., many states agreed on the Greenwich meridian to mark the conventional international zero meridian which is in use today.

At the beginning of the developments in satellite geodesy, the concept of geocentric Cartesian coordinates was found to be most convenient for all sorts of computations. It was, thus, only a small step to define the conventional global 3D coordinate system to have its origin at the geocenter, the z axis from the geocenter north along the rotation axis, the x axis pointing from the geocenter to the intersection of the equator and the zero meridian, and the y axis completing a right-hand system. This is of course only a theoretical definition. In the circles of the scientists dealing with this, this definition is called the *(reference) system*. In contrast to this, the *(reference) frame* is the physical realization of a system with markers such as reference points of radio telescopes plus the measured coordinates assigned to the markers. Please beware that the distinction of frame and system is not maintained everywhere.

In terms of the evolutionary steps of the terrestrial global reference frames, the story started with a number of astronomical observatories for which latitude, longitude and sometimes height information had been measured. The first space-geodesy techniques, satellite Doppler measurements and satellite laser ranging, adopted the latitude and longitude information by one way or another and converted them to global geocentric Cartesian coordinates for one or more reference instruments. Something similar was done for the first geodetic VLBI observations, mostly fixing the converted coordinates of one telescope and tagging the other telescopes to this reference through the observations.

As the first combined conventional global terrestrial reference frame, we can consider the BIH Terrestrial System 1984 (BTS84) of the Bureau International de l'Heure (BIH) which is the predecessor of the IERS (Boucher and Altamimi, 1990). The BTS84 incorporated 14 VLBI telescopes, 2 lunar laser ranging (LLR) telescopes and positions of 27 satellite laser ranging (SLR) telescopes, some of them mobile occupations. Since five of the sites had observations of both techniques and the geometric ties between them measured locally, it was possible to combine the technique-specific frames. At that time, station movements were not considered. This was only included years later in subsequent realizations of the conventional terrestrial reference frames [17]. The procedures used in the developments of terrestrial reference frames have to do a lot with datum definition which is addressed in more detail in Sec. 13.8.

Doing many evolutionary steps in one, today we make use of the International Terrestrial Reference System (ITRS) in its latest realization, the International Terrestrial Reference Frame 2020 (ITRF2020). Modern ITRFs combine observations and results of VLBI, SLR, GNSS and DORIS observing techniques. A combined conventional frame has the advantage that it makes use of the strengths of the individual techniques. For example, VLBI and SLR define the scale of the frame and in this respect control each other. In addition, and most importantly for VLBI, SLR defines the origin of the frame because low flying satellites are very sensitive to the gravity field of the Earth. Consequently, the geocenter is one of the primary products of SLR. In the ITRF computations, this information is transferred to the VLBI polyhedron.

Since the surface of the Earth is in motion due to tectonics and other geophysical phenomena including earthquakes, coordinates alone are not sufficient to describe the contemporary position of a reference marker. For this reason, ITRF2020 takes into account linear motions, discontinuities (earthquakes, man-made movements), post-seismic deformation or relaxations (exponential), and periodic inter-annual variations (annual and semi-annual). The full ITRF2020 model for telescope coordinates reads

$$\mathbf{x}_t = \mathbf{x}_{t_0} + \mathbf{v}_x \cdot (t - t_0) + \mathbf{\Delta} x_f + \mathbf{\Delta} x_{PSD} \tag{13.1}$$

with

$$\mathbf{\Delta} x_f = \sum_{i=1}^{n_f} (a' \cos(\omega_i t) + b' \sin(\omega_i t)) \tag{13.2}$$

for the contributions of the inter-annual variations $\mathbf{\Delta} x_f$, where a' and b' are the coefficients

[17]https://www.iers.org/IERS/EN/Publications/TechnicalNotes/TechnicalNotes.html

of the annual and semi-annual periods ω_i[18], and

$$\Delta x_{PSD} = \sum_{i=1}^{n^l} A_i^l \log(1 + \frac{t - t_i^l}{\mathcal{T}_i^l}) + \sum_{i=1}^{n^e} A_i^e \log(1 - e^{\frac{t-t_i^e}{\mathcal{T}_i^e}}) \tag{13.3}$$

for the post-seismic deformation Δx_{PSD} where n^l = Number of logarithmic terms of the parametric model, n^e = Number of exponential terms of the parametric model, A_i^l = Amplitude of the i^{th} logarithmic term, A_i^e = Amplitude of the i^{th} exponential term, $\mathcal{T}_i^l$ = Relaxation time of the i^{th} logarithmic term, $\mathcal{T}_i^e$ = Relaxation time of the i^{th} exponential term, t_i^l = Earthquake time(date) corresponding to i^{th} logarithmic term, t_i^e = Earthquake time(date) corresponding to the i^{th} exponential term.[19]

The coordinates at the reference epochs and the velocities are tabulated on the ITRF web site[20] (see also extract in Tab. 13.1.1). For sites with discontinuities, separate entries for the periods before and after the event are provided with different SOLN numbers (specific *solutions* for different periods) followed by the time section of their validity. DOMES numbers are special site identifiers established for ITRF work[21]. The full list of coefficients of Eq. 13.1 is available under https://itrf.ign.fr/en/solutions/ITRF2020.

Table 13.1: Extract from ITRF table ITRF2020_VLBI.SSC.txt

ITRF2020 STATION POSITIONS AT EPOCH 2015.0 AND VELOCITIES
VLBI STATIONS

DOMES NB.	SITE NAME	TECH.	ID.	X/Vx	Y/Vy	Z/Vz	Sigmas			SOLN	DATA_START	DATA_END
						m/m/y						
14209S001	EFFELSBERG	VLBI	7203	4033947.2868	486990.8233	4900431.0839	0.0092	0.0038	0.0105	1	00:000:00000	96:275:00000
14209S001				-.01397	0.01699	0.01070	.00034	.00016	.00040			
14209S001	EFFELSBERG	VLBI	7203	4033947.2977	486990.8298	4900431.0949	0.0047	0.0020	0.0056	2	96:275:00000	00:000:00000
14209S001				-.01397	0.01699	0.01070	.00034	.00016	.00040			

In Tab. 13.1.1, Effelsberg has two entries because in 1996 there was a man-made shift in the location of the reference point due to track repairs. To compute the x coordinates of Effelsberg at the epoch 2023.0 (we consider only the linear term), we can do so with

$$x_{2023.0} = 4033947.2977 - 0.01397 \cdot (2023.0 - 2015.0) \tag{13.4}$$

It should also be mentioned that in the framework of the International Earth Rotation and Reference Systems Service (IERS), further ITRS Product Centers exist which also produce global reference frames such as the DGFI Terrestrial Reference Frame (DTRF) and the JPL Terrestrial Reference Frame (JTRF). Both groups use techniques entirely different to that of the ITRF Product Center, and naturally come to very similar results although differences

[18] The choice of annual and semi-annual periods is in fact a Taylor's expansion, which is limited to the first two terms.
[19] https://itrf.ign.fr/ftp/pub/itrf/itrf2020/ITRF2020-PSD-model-eqs-IGN.pdf
[20] https://itrf.ign.fr/ftp/pub/itrf/itrf2020/ITRF2020_VLBI.SSC.txt
[21] https://itrf.ign.fr/en/network/domes/description

are sometimes significant.

For practical purposes most VLBI analysis groups maintain their own realizations from their latest global analyses. All of them are very close to the ITRF in its latest version (e.g., Altamimi et al., 2016) to which the IVS contributes routinely as described by Vennebusch et al. (2007), Böckmann et al. (2010), and Bachmann et al. (2016). On average the differences are only on the order of a few millimeters for the coordinates and some sub-millimeter per year differences for the velocities with very few exceptions for telescopes with known peculiarities or very short time histories.

13.1.2. Celestial reference frames

The second category of reference frames which is directly linked to VLBI observations consists of celestial frames related to the positions of the compact extra-galactic radio nuclei such as quasi-stellar objects (quasars) described in detail in Sec. 4. The positions of the objects are defined as angular right ascension α and declination δ components of an equatorial polar coordinate system (Fig. 4.3).

The conventional celestial reference frames tabulated in radio source catalogs, such as the International Celestial Reference Frames (ICRF) as realizations of the International Celestial Reference System (ICRS), have their origin in the solar system barycenter. For this reason, they realize a Barycentric Celestial Reference System (BCRS). This is close enough to the geocenter if light deflection and annual parallax are dealt with in the relativistic VLBI formulation (See Eq. 13.45). For this reason, the positions can also be considered to be valid in the Geocentric Celestial Reference System (GCRS).

Since the global rotation (of the sky before Galileo Galilei) produce a rotation axis with two poles in the sky, declinations, which are the latitudes in the sky, have been used to interpolate any position between them. Right ascension, which is the longitude in the sky, had to be defined by convention. However, other than on Earth, the sky had two distinguished points, the intersections of the equator and the ecliptic, called vernal equinox and autumnal equinox. The vernal equinox is the zero point of right ascension (R.A.) is counted positive in the direction of the Earth's spin in the equatorial plane from 0 to 24 hours with minutes and seconds of time (Fig. 4.3). Declination runs in degrees, minutes, and seconds of arc from -90° at the celestial south pole to +90° at the celestial north pole.

In the beginning of astrometry, the vernal equinox γ defined the zero meridian for positioning stars and other astronomical objects in the sky. However, today it is the other way round. The catalogue of the most recent ICRF through the positions of its numerous radio sources indirectly determines the location of the zero meridian. Since quasars and other compact radio sources do not exhibit any proper motions beyond apparent movements through source structure and core shift effects, the celestial reference frame serves as the reference for describing the movements of the vernal equinox on the celestial equator (see Sec. 13.2).

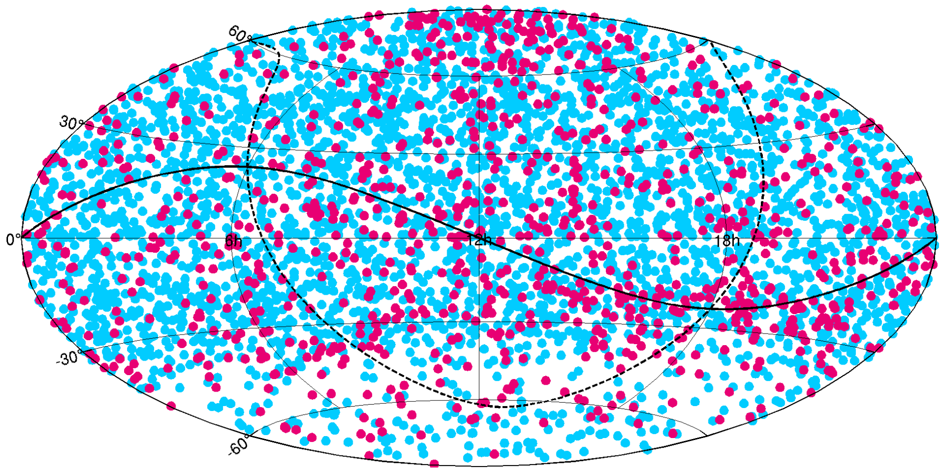


Figure 13.1: Source distribution of ICRF2 sources (blue) plus new sources in ICRF3 (magenta). The black line is the ecliptic. The dashed line follows the trace of the milky way (our galaxy).

The developments of the celestial reference frames have gone through similar stages as the terrestrial ones. Without going into too much detail of the past, the positions of the Fundamental Catalogue FK5 of optical identifications (Fricke et al., 1988) can be considered as the basis for those of modern radio frequency astrometry. Since the bright radio source 3C273B had an unresolved counterpart in the optical, its position was taken over from FK5 for the first VLBI analyses, sometimes as the only anchor point for the R.A. fixing. In 1995, the first conventional celestial reference frame in the radio frequency domain was computed as the International Celestial Reference Frame (Ma and Feissel, 1997). The catalogue of positions was aligned to FK5 within the uncertainty of the latter through preceding realizations compiled as the so-called ICRS (IERS Celestial Reference System) (Arias et al., 1995).

For quite some time the most accurate frame had been the International Celestial Reference Frame in its second realization (ICRF2) (Fey et al., 2015). However, in 2019 work has been completed to make use of an additional wealth of observations, which has been accumulated between March 2009 and October 2017, and generate a new ICRF, i.e., ICRF3 (Charlot et al., 2020). The axis definition is accurate to 10–20 μas and the position uncertainties have a median error of 0.15 mas as compared to 0.5 mas in ICRF2. For further details, please read de Witt et al., 2022.

It should be mentioned that now there is a new very precise catalog in the optical domain originating from the Gaia mission. However, this catalogue is not directly accessible by radio telescopes but elaborate studies are underway to identify counterparts in both frequency domains.

13.2. Earth orientation parameters

The Earth exhibits a rather variable rotation which is caused by a number of physical causes (Lambeck, 1980). There are multiple ways of formulating Earth rotation mainly separating dynamic and kinematic aspects. For our VLBI purposes, we will concentrate on the kinematic formulations.

For describing the instantaneous orientation of the terrestrial reference frame in the celestial reference frame at a certain epoch, and vice versa, it is sufficient to apply three rotation angles. These can be Euler angles, which rotate around one axis first, then around a second axis, and last around the first axis again. With Cardanian angles the rotations are applied to all three axes one after another. The timely evolution of these angles is difficult to interpret. However, the early astrometric observations were carried out from the surface of the rotating Earth, first detecting only precession (Hipparcos, 135 B.C.), then later also nutation (Bradley, 1728), and, thus, made use of the Earth's rotation axis. Although precession and nutation are caused by the same phenomenon, torques onto the equatorial bulge of the Earth and its response for angular momentum conservation, they have always been separated because of their entirely different periods of $\sim$26,000 of precession and 18.6 years and below of nutation.

That Earth's rotation axis is not fixed with respect to the body of the Earth was known theoretically since the days of Leonhard Euler (1707–1783) but became obvious only with the measurements of the International Latitude Service starting in 1899. Only with the concept of an axis of rotation and its localization through its poles on related spheres, the Earth's variable rotation can be described by five angular parameters, two components for the celestial pole coordinates describing the movements of the rotation axis in space attributed to precession and nutation, the Earth's phase of rotation and two components describing the movements of the rotation axis with respect to the Earth's crust, called polar motion.

The theory of the Earth's variable rotations and their modelling are being worked on almost constantly because there still are many unanswered questions. Concerning the kinematic description, we have seen a major change in the concept, mostly due to improved observing capabilities and advanced modelling of geophysical phenomena. An important argument for the change was also the separation of the celestial reference system from the Earth's variable rotation. In the following, we will only look at the new paradigm which is in agreement with the resolutions of the International Astronomical Union (IAU) of 2000 and 2006. These resolutions in particular render obsolete the old definitions of Universal Time (UT1) and the precession/nutation formulation related to the vernal equinox.

The IAU 2000/2006 paradigm hinges on the introduction of so-called non-rotating origins (NRO) which were proposed by Guinot (1979) and introduced by Capitaine et al. (1986), and Capitaine et al. (2000). The Celestial Intermediate Origin (CIO) is the non-rotating origin on the moving equator of the Celestial Intermediate Pole (CIP) (Fig. 13.3).

The Terrestrial Intermediate Origin (TIO) is the non-rotating origin on the Earth's equator. These intermediate origins, which strictly speaking are rather zero-meridians, had been called ephemeris origins in the above references. These expressions were changed by a resolution of the International Astronomical Union (IAU) in 2006 (B2).

For performing the necessary transformations needed for modeling the a priori delays and also for later EOP estimation, we should first understand the sequence of necessary rotations. The transformation of a point with coordinates in the International Terrestrial Reference System (ITRS) into those in the Geocentric Celestial Reference System (GCRS) is realized by sequentially applying the basic rotations for polar motion (wobble), Earth's spin as well as nutation and precession. These are applied through the rotation matrices **W**, **S** and **Q**, respectively, where **Q** comprises both precession and nutation (Eq. 13.5). The sequence here is inverted so that the rotations are applied to the ITRS coordinates "from the left" because the axes are always rotated to new intermediate directions.

$$[GCRS] = \mathbf{Q} \cdot \mathbf{S} \cdot \mathbf{W} \cdot [ITRS] \tag{13.5}$$

The sequence of rotations is important for the signs of the arguments. For details on the composition of the fundamental rotation matrices, please see Appendix A.

13.2.1. Polar motion

The variability of the Earth's axis of rotation with respect to the Earth's crust is parameterized by the two components x_p and y_p of polar motion contained in the rotation matrix of this wobble effect $\mathbf{W}'(x_p(t), y_p(t))$[22]. For historical reasons, the polar motion x component is positive in the same direction as the geocentric Cartesian x axis while the y component is positive towards the 270° meridian, i.e., positive to the West and, thus, negative towards the geocentric Cartesian y axis (Fig. 13.2) . Consequently, the coordinates of the CIP pole in the TRF are x_p and $-y_p$. This is quite convenient because, as we see from Fig. 13.2, the rotations needed for the transition of the CIP in the *TRF of convention*, e.g., ITRF, to the *TRF of date* are both positive (counter-clockwise) requiring the rotations $\mathbf{R}_2(x_p)$ and $\mathbf{R}_1(y_p)$ according to Eqs. A.8 and A.7 in Appendix A[23] . The question now arises which of the two very small rotations have to be applied first. Putting the $\mathbf{R}_1$ rotation first, the multiplication with $\mathbf{R}_2$ (from the left) yields

$$\mathbf{R}_2(x_p) \cdot \mathbf{R}_1(y_p) = \begin{pmatrix} \cos x_p & \sin x_p \sin y_p & -\sin x_p \cos y_p \\ 0 & \cos y_p & \sin y_p \\ \sin x_p & -\cos x_p \sin y_p & \cos x_p \cos y_p \end{pmatrix} \tag{13.6}$$

[22]The prime is used to indicate that the wobble matrix is still incomplete.
[23]R_1 around axis 1 and $R_2 around axis 2$

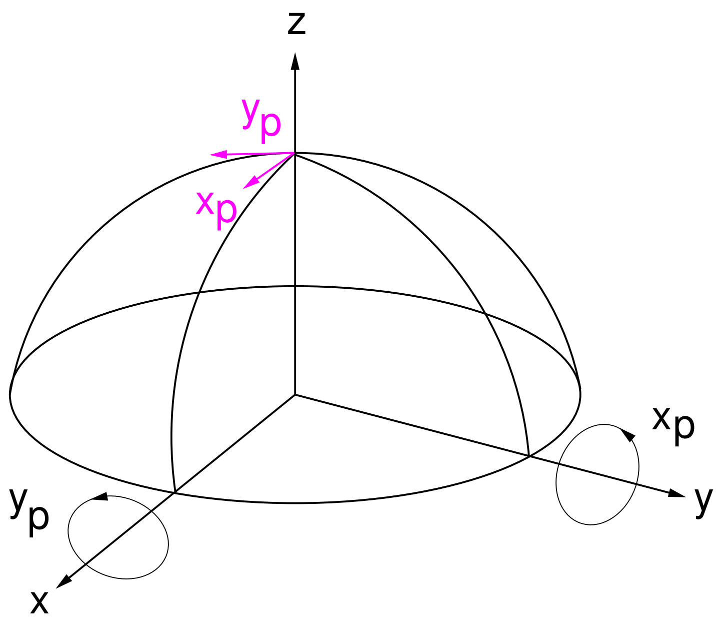


Figure 13.2: Rotations for polar motion. The sphere is supposed to represent the surface of the Earth with the arc in the X/Z plane standing for the prime or zero meridian.

Simplifying the elements of this rotation matrix for the very small polar motion angles of up to 600 mas (3×10^{-6} radians) with $\cos x_p/y_p \approx 1$ and $\sin x_p/y_p \approx x_p/y_p$, we find

$$\mathbf{R}_2(x_p) \cdot \mathbf{R}_1(y_p) = \begin{pmatrix} 1 & 0 & -x_p \\ 0 & 1 & y_p \\ x_p & -y_p & 1 \end{pmatrix} \tag{13.7}$$

Swapping the order of the matrix multiplication, we get the same result, so

$$\mathbf{R}_2(x_p) \cdot \mathbf{R}_1(y_p) \approx \mathbf{R}_1(y_p) \cdot \mathbf{R}_2(x_p) \tag{13.8}$$

In the literature, mostly the first type of ordering is used, so we can write for the rotations of polar motion

$$\mathbf{W}'(t) = \mathbf{R}_2(x_p) \cdot \mathbf{R}_1(y_p). \tag{13.9}$$

The wobble matrix $\mathbf{W}'$ is only part of the rotations for polar motion because the TIO is separated from the conventional prime meridian by a small angle $s'(t)$, which is the TIO locator (Terrestrial Intermediate Origin) (Dehant and Mathews, 2015). The entire rotation matrix for the transformation of the CIP in the *TRF of convention* to the *TRF of date* is then composed of

$$\mathbf{W}(t) = \mathbf{R}_3(-s') \cdot \mathbf{R}_2(x_p) \cdot \mathbf{R}_1(y_p). \tag{13.10}$$

13.2.2. Earth's phase of rotation

With the change in paradigm, the Earth's phase of rotation is measured as the so-called *Earth rotation angle* θ between the Celestial Intermediate Origin (CIO) and the Terrestrial Intermediate Origin (TIO) (Capitaine et al., 2000)(Fig. 13.3). It should be noted here that originally this was called the *stellar angle*. To make the *Earth rotation angle* θ a time quantity, a conventional relationship was defined as

$$\theta(T_u) = 2\pi \cdot (0.7790572732640 + 1.00273781191135448 \cdot (T_u - 2451545.0)) \quad (13.11)$$

with T_u = Julian UT1 date (Capitaine et al. (2000), IAU Resolution 2000 B1.8). According to the convention laid out in Appendix A (Eq. A.9), the rotation matrix of the Earth's phase of rotation is

$$\mathbf{R}_3(\theta) = \mathbf{S}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (13.12)$$

However, the argument θ needs to be applied with a negative sign in the application to Eq. 13.5 because it is increasing due to the Earth's spin, i.e., with a positive sign, and needs to be counter-acted in the transformation. Applying the rotation matrix (Eq. 13.12) to a vector of terrestrial coordinates *of date* is equivalent to rotating these from the TIO system *of date* into the CIO system *of date*.

Considering the application of Eq. 13.11 in the opposite direction provides UT1 from *Earth rotation angles* θ measured by VLBI observations (in fact the only means to measure UT1). Solving for T_u in Eq. 13.11 yields

$$UT1 = \frac{\theta(T_u)/2\pi - 0.7790572732640}{1.00273781191135448} + 2451545.0 - JD \quad (13.13)$$

JD is the Julian Day of the observation epoch in decimal fractions of the day.

13.2.3. Precession and nutation

The two components of the celestial pole coordinates are used to model the motion of the Earth's axis of rotation in space, generally called precession and nutation. Its variability is mainly caused by luni-solar torques on the equatorial bulge of the Earth and the conservation of angular momentum which forces the rotation axis to shift and nod. Although precession and nutation are actually caused by the same physical phenomena with two very distinct time scales (~26,000 years vs. 18.6 years and down to a few days) they had been separated for a long time because of the time scales and because precession was detected much earlier than nutation (135 B.C. vs. 1728).

Today, the kinematic description treats the two together with the angular arguments being the two angular values X and Y (Fig. 13.3). These are used to describe the Celestial

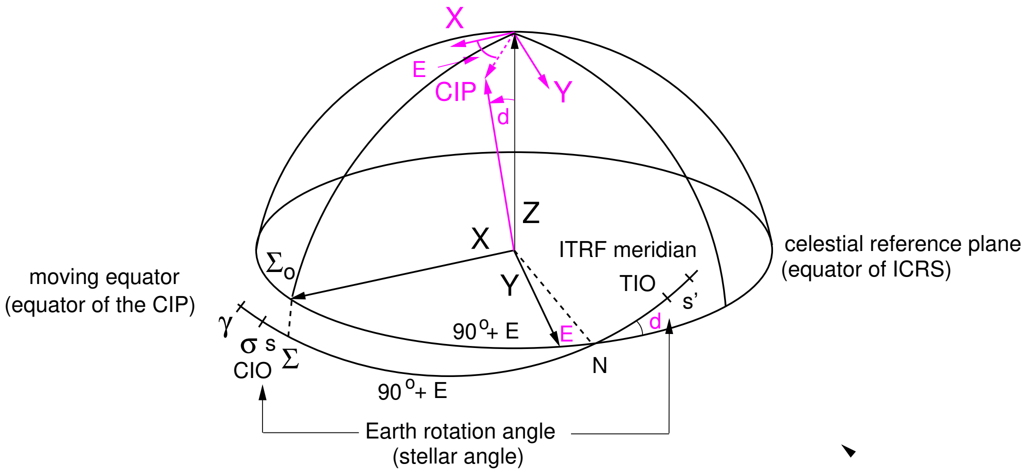


Figure 13.3: Rotations due to the Earth rotation angle (stellar angle) and precession and nutation. The sphere is supposed to represent the celestial sphere with global X,Y,Z coordinates in the Geocentric Celestial Reference Frame.

Intermediate Pole (CIP) in the celestial frame (Capitaine, 1990). However, conceptually the derivation of the numerical model starts with the angles d and E which are closely related to the angles z_A, ζ_A and Θ_A of the old, equinox-based theory of precession and nutation (Capitaine, 1990). With d and E, the rotation matrix $\mathbf{Q}'(E,d)$ relates the CIP to the origin of the celestial reference system Σ_0. Formulating the rotations from the GCRS to the instantaneous frame (frame of date), Capitaine (1990) develops this rotation matrix as:

$$\mathbf{Q}'(E,d) = \mathbf{R}_3(-(90^\circ + E)) \cdot \mathbf{R}_1(d) \cdot \mathbf{R}_3(90^\circ + E). \tag{13.14}$$

It is an Eulerian rotation sequence following the scheme of the first rotation around the first axis, then around a second axis and in the opposite direction around the first axis again. Eq. 13.14 is generally simplified through an elimination of the two back and forth 90° rotations and the rotation around the 2-axis instead of around the 1-axis (Lieske, 1979):

$$\mathbf{Q}'(E,d) = \mathbf{R}_3(-E) \cdot \mathbf{R}_2(d) \cdot \mathbf{R}_3(E). \tag{13.15}$$

Populating the rotation matrices with the respective angles and multiplying them yields

$$\mathbf{Q}'(t)(E,d) = \begin{pmatrix} 1+\cos^2 E(\cos d - 1) & (\cos d - 1)\sin E \cos E & -\sin d \cos E \\ (\cos d - 1)\sin E \cos E & 1+\sin^2 E(\cos d - 1) & -\sin d \sin E \\ \sin d \cos E & \sin d \sin E & \cos d \end{pmatrix}. \tag{13.16}$$

However, numerical values for d and E have never been derived directly. One of the reasons is the singularity of E near J2000.0 as can be seen in Fig. 13.5. d is always positive as can be seen in Fig. 13.4. Nonetheless, the rotation matrix $Q'(d,E)$ could be used to describe the position of the CIP in the celestial reference frame if needed.

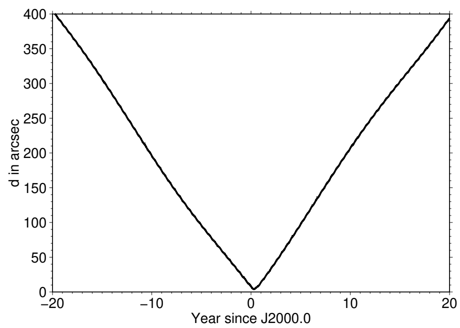


Figure 13.4: Angle d with singularity near J2000.0 (computed indirectly through the inverse of Eq. 13.17).

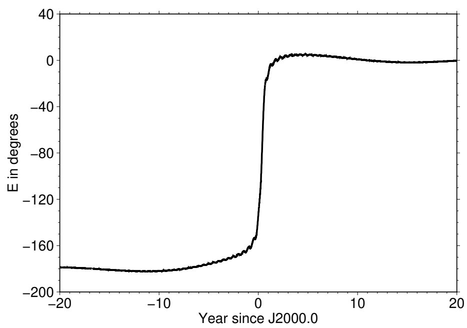


Figure 13.5: Angle E with singularity near J2000.0 (computed indirectly through the inverse of Eq. 13.17).

Instead, Capitaine (1990) started to express the position of the CIP in the celestial coordinates $X(t), Y(t)$ which are related to d and E through

$$X = \sin d \cos E, \qquad Y = \sin d \sin E, \qquad Z = \cos d. \tag{13.17}$$

X and Y are the projections of the unit vector in the direction of the CIP on the Cartesian celestial coordinate system (Capitaine, 1990). In this concept, d and E can be seen as spherical polar coordinates of the CIP. The units of X and Y are radians initially but can be translated into arcseconds easily.

In numerical terms, X and Y are computed (Petit and Luzum, 2010) by

$$\begin{aligned} X \quad = \quad & -0.016617" + 2004.191898"\, t - 0.4297829"\, t^2 \\ & -0.19861834"\, t^3 + 0.000007578"\, t^4 + 0.0000059285"\, t^5 \\ & + \sum_i [(a_{s,0})_i \sin(ARGUMENT) + (a_{c,0})_i \cos(ARGUMENT)] \\ & + \sum_i [(a_{s,1})_i\, t\, \sin(ARGUMENT) + (a_{c,1})_i\, t\, \cos(ARGUMENT)] \\ & + \sum_i [(a_{s,2})_i\, t^2\, \sin(ARGUMENT) + (a_{c,2})_i\, t^2\, \cos(ARGUMENT)] \\ & + \ldots, \end{aligned} \tag{13.18}$$

$$
\begin{aligned}
Y \;=\; & -0.006951" - 0.025896"\,t - 22.4072747"\,t^2 \\
& +0.00190059"\,t^3 + 0.001112526"\,t^4 + 0.0000001358"\,t^5 \\
& +\sum_i [(b_{s,0})_i \sin(ARGUMENT) + (b_{c,0})_i \cos(ARGUMENT)] \\
& +\sum_i [(b_{s,1})_i\, t \, \sin(ARGUMENT) + (b_{c,1})_i\, t \, \cos(ARGUMENT)] \\
& +\sum_i [(b_{s,2})_i\, t^2 \, \sin(ARGUMENT) + (b_{c,2})_i\, t^2 \, \cos(ARGUMENT)] \\
& +\ldots,
\end{aligned} \tag{13.19}
$$

It should be emphasized that also X and Y describe the position of the CIP in the celestial reference frame. The polynomial parts of the X, Y series originate from precession, except for the contribution from the frame bias and from cross nutation terms (Capitaine and Wallace, 2006). For more details on the periodic (nutation) terms please refer to Mathews et al. (2002) and Petit and Luzum (2010). As can be seen from Fig. 13.6 (and Fig. 13.7), where the X and Y values are computed according to Eqs. 13.19 and 13.20, the major progress is in the X component, which is along the precession cone. At J2000.0, X is -5.6" and Y is -5.8".

In the old equinox-based precession theory, the precession constant is generally quoted as 50.3"/y. It is measured along the ecliptic. The approximate relation between the "old" and "new" celestial pole offsets is $X \approx \Delta\psi \sin\varepsilon_0$ and $Y \approx \Delta\varepsilon$. So, if the precession constant (in $\Delta\psi$, measured along the ecliptic) is 50"/y, it becomes about 20"/y in X though the multiplication with the sine of the angle between the celestial equator and the ecliptic of $\varepsilon_0 \approx 23.5°$.

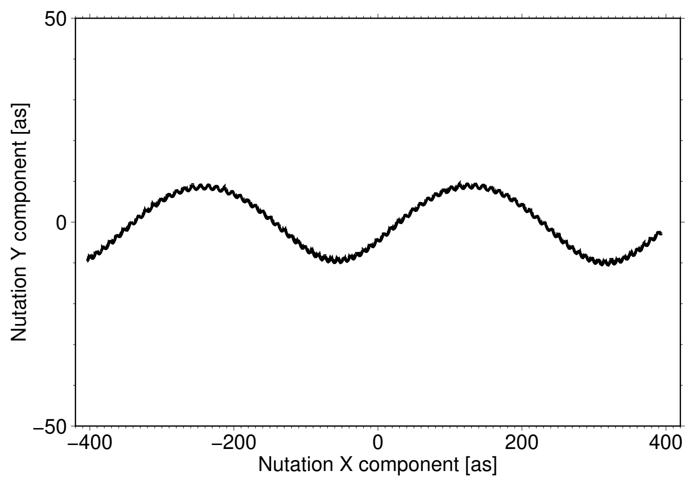


Figure 13.6: Precession and nutation in the X/Y projection between 1980 and 2020. The graph is dominated by the precession of about 20"/y and approximately two cycles of the 18.6 yr period.

From these X and Y values, the composite rotation matrix, $\mathbf{Q}'(t)(X,Y)$, can be con-

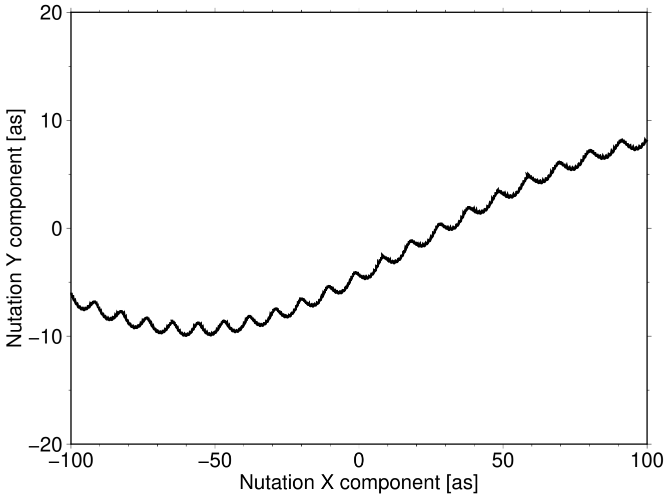


Figure 13.7: Precession and nutation in the X/Y projection between 1995 and 2005. The superimposed periodicity stems from the 182.6 day period.

structed (Capitaine, 1990; Capitaine and Wallace, 2006) directly as

$$\mathbf{Q}'(t)(X,Y) = \begin{pmatrix} 1-aX^2 & -aXY & -X \\ -aXY & 1-aY^2 & -Y \\ X & Y & 1-a(X^2+Y^2) \end{pmatrix} \tag{13.20}$$

with

$$a = \frac{1}{1+\cos d} = \frac{1}{2} + \frac{1}{8}(X^2+Y^2) \tag{13.21}$$

This rotation matrix (13.20) is still the one to describe the position of the CIP in the GCRS. For the coordinate transformation between *GCRS of date* to the *GCRS of convention* (J2000.0) we need to apply the transposed matrix

$$\mathbf{Q}'^T(t)(X,Y) = \begin{pmatrix} 1-aX^2 & -aXY & X \\ -aXY & 1-aY^2 & Y \\ -X & -Y & 1-a(X^2+Y^2) \end{pmatrix} \tag{13.22}$$

It can be used for applications with an accuracy requirement of 1 μas. For completeness, it should be mentioned that Eq. 13.22 corresponds to the counter-rotation of 13.15 which is

$$\mathbf{Q}(E,d) = \mathbf{R}_3(-E)\cdot\mathbf{R}_2(-d)\cdot\mathbf{R}_3(E). \tag{13.23}$$

In addition to the X/Y rotations, a further, though small, rotation with the angle $s(t)$ around the 3-axis is needed. $s(t)$ is the so-called CIO-Locator, which is the position of the

CIO in the GCRS (Fig. 13.3). It is computed according to

$$s(t) = -\frac{1}{2}[X(t)Y(t) - X(t_0)Y(t_0)] + \int_{t_0}^{t} \dot{X}(t)Y(t)\,dt - (\sigma_0 N_0 - \Sigma_0 N_0) \tag{13.24}$$

where σ_0 and Σ_0 are the positions of the Celestial Intermediate Origin (CIO) at J2000.0 and the x-origin of the GCRS respectively and N_0 is the ascending node of the equator at J2000.0 in the equator of the GCRS (Petit and Luzum, 2010). $(\sigma_0 N_0 - \Sigma_0 N_0)$ is the so-called Equation of the Origins (EO).

In numerical terms, s(t) can be derived according to Tab. 5.2d in Petit and Luzum (2010):

$$s(t) = \frac{-XY}{2} + 94 + 3808.65\,t - 122.68\,t^2 - 72574.11\,t^3 + ... \tag{13.25}$$

with t being t = (TT - 2000 January 1d 12h TT) in days/36525. The expression is followed by a number of small terms depending on fundamental arguments. The date 2000 January 1.5 TT = Julian Date 2451545.0 TT (Temps Terrestre) is the definition of J2000.0. For more details see the IERS Conventions 2010 (Petit and Luzum, 2010).

The full rotation model for precession and nutation for positions in the *GCRS of date* to the *GCRS of convention* (J2000.0) then consists of

$$\mathbf{Q}(X,Y,s) = \mathbf{Q}'^{T}(X,Y) \cdot \mathbf{R}_3(s) \tag{13.26}$$

13.2.4. Full Earth orientation parameter transformation

The transformation of a point with coordinates in the International Terrestrial Reference System (ITRS) into those in the Geocentric Celestial Reference System (GCRS) is realized by sequentially applying the rotation matrices introduced above (Petit and Luzum, 2010):

$$\mathbf{x}_{GCRS} = \mathbf{Q}'^{T}(X,Y) \cdot \mathbf{R}_3(s) \cdot \mathbf{S}(-\theta) \cdot \mathbf{R}_3(-s') \cdot \mathbf{W}'(x_p, y_p) \cdot \mathbf{x}_{ITRS} \tag{13.27}$$

For the transformation in the opposite direction, the sequence as well as the angles have to be inverted so that

$$\mathbf{x}_{ITRS} = \mathbf{W}(-y_p, -x_p) \cdot \mathbf{R}_3(s') \cdot \mathbf{S}(\theta) \cdot \mathbf{R}_3(-s) \cdot \mathbf{Q}(-X, -Y) \cdot \mathbf{x}_{GCRS} \tag{13.28}$$

13.3. Geodetic and astrometric VLBI functional model

Data analysis of geodetic and astrometric VLBI analysis follows the same rules as any other geodetic problem with a surplus of observations for the determination of the parameters of interest. Let's first look at the purely geometric relationships following the principal VLBI equation (Eq. 2.7) which has to be expanded to take into account the fact that the baseline vector **b** and the unit vector in source direction **k** have to be expressed in the same reference frame. The main cause for a more elaborate formulation is the Earth's variable rotation (Lambeck, 1980). For this reason, the rotation matrices for precession and nutation $\mathbf{Q}(X(t), Y(t))$, daily spin $\mathbf{S}(\theta(t))$, and polar motion $\mathbf{W}(x_p(t), y_p(t))$ according to Sec. 13.2 have to be applied.

Application of these three rotation matrices either transforms the baseline vector in the terrestrial reference frame into that in the celestial reference frame or transforms the unit vector in the celestial reference into that in the terrestrial frame. Both transformations are equivalent and can also be applied in part to the radio sources and in part to the baseline vector, as is done in practice (see below). In a closed formula, they are embedded in

$$\tau(t) = T_{(2)} - T_{(1)} = -\frac{1}{c}\,\mathbf{b} \cdot \mathbf{W}(t) \cdot \mathbf{S}(t) \cdot \mathbf{Q}(t) \cdot \mathbf{k} \tag{13.29}$$

with

$$\mathbf{b} = \begin{pmatrix} x_{(2)} - x_{(1)} \\ y_{(2)} - y_{(1)} \\ z_{(2)} - z_{(1)} \end{pmatrix} \tag{13.30}$$

for the Cartesian coordinates of telescopes 1 and 2, and

$$\mathbf{k} = \begin{pmatrix} \cos\delta_c \cdot \cos\alpha_c \\ \cos\delta_c \cdot \sin\alpha_c \\ \sin\delta_c \end{pmatrix} \tag{13.31}$$

for the position of the quasars in the celestial system (Sec. 13.1.2).

For the full expansion of the observed delay, we have to look at the relativistic contributions to the functional model and at several further effects to consider. Depending on the nature of these effects they are either applied as delay corrections $\Delta\tau$ or as corrections to the coordinate vector of the radio telescope $\mathbf{x}_i$. It should be noted that at first we only consider those contributions, which can be applied a priori, while under Sec. 13.7.2 we address remaining model components, which can only be estimated.

13.4. Relativistic VLBI model delay

As we will see later (Sec. 13.7.2), we need to compute a very accurate a priori delay from all the information we have in advance. The main reason is that the least-squares adjustment will work with linearized quantities, such as partial derivatives and reduced observations, meaning that we limit the adjustment process to the last very few decimeter of an otherwise thousands of kilometer large baseline to radio source geometry.

At this point relativistic effects need to be taken into account for proper modeling of the delays. The first item to consider is the transformation of time scales. We have to distinguish proper time measured at an arbitrary location and coordinate time which refers to the barycenter of a coordinate system. Any clock on Earth, first of all, measures proper time, while we will need time as coordinate time of the solar system barycenter. The proper time at a telescope is normally given in the UTC time scale (Universal Time Coordinated) which relates to the International Atomic Time TAI (Temps Atomique International) through an integer number of leap seconds T_{LS} to keep UTC within 0.9 s of the astronomical time UT1. Terrestrial Time (TT) is an ideal time which a clock would indicate on the geoid. It has the same rate as TAI but an offset of 32.184 s originating from a changeover from Ephemeris Time (ET) to TAI, thus

$$TT = TAI + 32.184\,s = UTC + T_{LS} + 32.184\,s. \tag{13.32}$$

The coordinate time of the Earth is Temps Coordonnées Géocentric (TCG) or Geocentric Coordinate Time. TCG [in fractions of Julian date] relates to TT [in fractions of Julian date] (Bangert et al., 2017) through

$$TCG = TT + L_G \cdot (T_{JD} - TT_0) \tag{13.33}$$

with T_{JD} being TT expressed as Julian date, TT_0 being 2443144.5003725 (i.e., TT at 1977 January 1.0 TAI), and $L_G = 6.969290134 \times 10^{-10}$ (Petit and Luzum, 2010).

The coordinate time of the solar system barycenter TCB (Temps Coordonnées Barycentrique) is related to TCG through

$$TCB - TCG = \frac{1}{c^2}\left(\int_{t_0}^{t}[\frac{v_\oplus^2}{2} + U_{ext}(\mathbf{X}_\oplus)]dt + \mathbf{V}_\oplus \cdot (\mathbf{X} - \mathbf{X}_\oplus)\right) \tag{13.34}$$

where the vectors $\mathbf{X}_\oplus$ and $\mathbf{V}_\oplus$ denote the barycentric position and velocity of the geocenter. The vector $\mathbf{X}$ is the barycentric position of the observer and U_{ext} is the Newtonian potential of all of the solar system bodies apart from the Earth, evaluated at the geocenter (Bangert et al., 2017). t is TCB and t_0 is 1977 January 1.0 TAI.

Now we turn to the differences in arrival times at the two radio telescopes. The theoretical time difference d_{TT} can be considered as being the same as that of d_{UTC}

$$d_{TT} = d_{TAI} = d_{UTC} = (t_{(2)})_{TT} - (t_{(1)})_{TT} \tag{13.35}$$

From the proper time frame to the Geocentric Coordinate Time (TCG) frame, the time difference needs to be scaled by

$$d_{TCG} = \frac{d_{TT}}{1 - L_G} \tag{13.36}$$

with $L_G = 6.969290134 \times 10^{-10}$ as above. It should be noted here that the VLBI coordinate frame is in fact (only) related to the TT time scale ("TT-compatible") for consistency with the other observing techniques in the ITRF as decided in an ITRF workshop in 2000, although IAU and IUGG resolutions call for TCG-compatibility (Petit and Luzum, 2010).

The next step is the transformation of the Earth-fixed geocentric coordinates of the radio telescopes, e.g., given in the ITRS (Sec. 13.1.1), $\mathbf{x}_i$ into the barycentric frame $\mathbf{X}_i$. For this,

$\mathbf{x}_i$ has to be transformed from ITRS into the Geocentric Celestial Reference System (GCRS) which has the same origin but fixed axis directions with respect to the extra-galactic radio sources of the ICRS. The transformation then just consists of adding the barycentric radius vector of the geocenter $\mathbf{X}_{\oplus}(t_1)$ to the radius vector of telescope i $\mathbf{x}_i$ (in GCRS).

$$\mathbf{X}_i(t_{(1)}) = \mathbf{X}_{\oplus}(t_{(1)}) + \mathbf{x}_i(t_{(1)}) \tag{13.37}$$

The reference epoch is $t_{(1)}$, i.e., when the signal arrived at telescope (1). All quantities to be addressed further down will also be referred to this time epoch because this epoch is the reference time of observation. Here it is in the TCG time scale which needs a scaling as in Eq. 13.33.

A very important contribution to the VLBI time delay originates from the retardation of the signal paths through gravitational bending of the celestial bodies. It is part of the relativistic formulation of the delay (Eq. 13.45). For modeling the bending effect, the time of closest approach of a signal originating from a quasar and passing a celestial body is needed. For the J^{th} gravitating body, it can be expressed as the minimum

$$t_{1J} = \min\left[t_{(1)}, t_{(1)} - \frac{\mathbf{K} \cdot (\mathbf{X}_J(t_{(1)}) - \mathbf{X}_{(1)}(t_{(1)}))}{c} \right]. \tag{13.38}$$

$\mathbf{K}$ is the unit vector from the solar system barycenter to the source in the absence of gravitational or aberrational bending (Petit and Luzum, 2010). This minimum actually means that it stays at $t_{(1)}$ if the signal does not pass the body before it arrives at telescope (1).

For the final expression, we need the vectors between the positions of the two telescopes and the gravitating body at the time of nearest approach, $\mathbf{R}_{1J}(t_1)$ and $\mathbf{R}_{2J}(t_1)$:

$$\mathbf{R}_{1J}(t_{(1)}) = \mathbf{X}_{(1)}(t_{(1)}) - \mathbf{X}_J(t_{1J}) \tag{13.39}$$

$$\mathbf{R}_{2J}(t_{(1)}) = \mathbf{X}_{(2)}(t_{(1)}) - \frac{\mathbf{V}_{\oplus}}{c}(\mathbf{K} \cdot \mathbf{b}) - \mathbf{X}_J(t_{1J}) \tag{13.40}$$

$\mathbf{V}_{\oplus}$ is the barycentric velocity of the geocenter and $\mathbf{b}$ is the geocentric baseline vector.

The general relativistic delay, ΔT_{grav}, a.k.a. Shapiro term, then is

$$\Delta T^J_{grav} = (1+\gamma)\frac{GM_J}{c^3}\ln\frac{|\mathbf{R}_{1J}| + \mathbf{K} \cdot \mathbf{R}_{1J}}{|\mathbf{R}_{2J}| + \mathbf{K} \cdot \mathbf{R}_{2J}} \tag{13.41}$$

with M_J being the rest mass of the J^{th} gravitating body and γ the so-called light deflection parameter equal to 1 according to the Theory of General Relativity (GRT).

The gravitational delay due to the Earth at the picosecond level is

$$\Delta T^{\oplus}_{grav} = (1+\gamma)\frac{GM_{\oplus}}{c^3} ln\frac{|\mathbf{x}_{(1)}| + \mathbf{K} \cdot \mathbf{x}_1}{|\mathbf{x}_{(2)}| + \mathbf{K} \cdot \mathbf{x}_{(2)}} \tag{13.42}$$

with $M_\oplus$ being the rest mass of the Earth. The total gravitational delay can be expressed as the sum over all gravitating bodies, i.e., the Sun, Earth, the Earth's moon and the other planets,

$$\Delta T_{grav} = \sum_J \Delta T^J_{grav} \tag{13.43}$$

In the barycentric frame, the vacuum delay is

$$T_{(2)} - T_{(1)} = -\frac{1}{c}\mathbf{K} \cdot (\mathbf{X}_{(2)}(T_{(2)}) - \mathbf{X}_{(1)}(T_{(1)})) + \Delta T_{grav} \tag{13.44}$$

Eq. 13.44 is now converted (back) by a Lorentz transformation from the barycentric into the geocentric system. This is done w.r.t. two groups of quantities, the barycentric vectors $\mathbf{X}_i$ into the corresponding geocentric vectors $\mathbf{x}_i$ and the related transformations between the barycentric time difference ($T_{(2)}$ - $T_{(1)}$) and the geocentric time difference ($t_{(2)}$ - $t_{(1)}$).

The geometric delay then follows

$$(t_{(2)} - t_{(1)})_{vac} = \frac{\Delta T_{grav} - \dfrac{\mathbf{K}\cdot\mathbf{b}}{c}\left[1 - \dfrac{(1+\gamma)U}{c^2} - \dfrac{|\mathbf{V}_\oplus|^2}{2c^2} - \dfrac{\mathbf{V}_\oplus\mathbf{w}_{(2)}}{c^2}\right] - \dfrac{\mathbf{V}_\oplus\mathbf{b}}{c^2}\left(1 + \dfrac{\mathbf{K}\mathbf{V}_\oplus}{2c}\right)}{1 + \dfrac{\mathbf{K}(\mathbf{V}_\oplus + \mathbf{w}_{(2)})}{c}}. \tag{13.45}$$

This is the formulation for the geometry in vacuum (Petit and Luzum, 2010). $\mathbf{w}_{(i)}$ is the geocentric velocity vector of telescope (i), i.e., w.r.t. the geocenter. $\mathbf{V}_\oplus$ contains the x-, y-, z-velocity components of the geocenter itself. U is the total gravitational potential of the solar system at the geocenter neglecting the effects of the Earth's mass. This formula thus contains the effects of daily and annual aberration. In the context of VLBI, this means that while the signal continues traveling for a time τ after arriving at telescope (1), telescope (2) has changed its position w.r.t. that at $t_{(1)}$ both in the geocentric and in the solar system barycentric frame.

For completeness, we should briefly look at the evolution of this relativistic model. The formulations in the IERS Conventions (2010) (Petit and Luzum, 2010) originate from a joint effort by a group of scientists led by Marshall Eubanks resulting in the "Consensus Model" (Eubanks, 1991). The model appeared in the IERS Technical Note series[24] in 1992 as part of the IERS Standards (1992) (McCarthy, 1992) and was then carried over to the IERS Conventions (1996) (McCarthy, 1996) and IERS Conventions (2003) (McCarthy and Luzum, 2004).

13.5. Atmospheric refraction

Another very important contribution to the delay is that of atmospheric refraction causing a bending of the signal path and leading to a change in velocity of the signal's propagation. The latter mainly causes a retardation but has an accelerating effect for phase observations.

[24] https://www.iers.org/IERS/EN/Publications/TechnicalNotes/TechnicalNotes.html

In terms of refraction, the atmosphere is separated into the charged compartment, generally called the ionosphere, and the neutral part which some authors call the troposphere. This, however, falls short of the fact that only about 75% of the refraction effects in the neutral atmosphere take place within the troposphere while 25% happen above (Elgered, 1993).

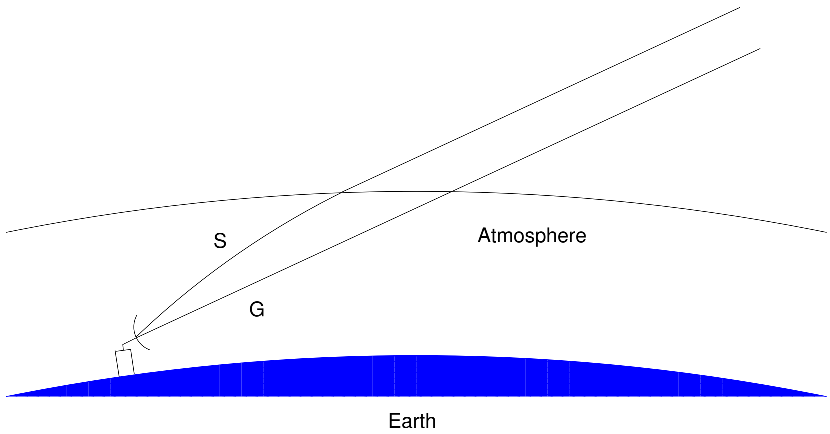


Figure 13.8: Refraction geometry. G = path in vacuum (straight line), S = geometrical path of electromagnetic wave (Elgered, 1993).

The path of a signal in vacuum is, apart from relativistic effects, a straight line G (Fig. 13.8). In contrast, the path of the electromagnetic wave in the atmosphere is curved with the electrical path length L along the geometric path length S depending on the refraction coefficient n

$$L = \int_S n\, ds \tag{13.46}$$

where n in turn depends on the velocity of light and the signal's velocity v with

$$n = \frac{c}{v} \tag{13.47}$$

According to Fermat's principle, the radiation as seen from the radio telescope attempts to leave the layer with a higher refraction index (lower layer) towards a lower refraction index (higher layer), thus bending the path upwards. This concept minimizes the electrical path length L but is valid only for stratified layers. As soon as the curvature of the atmosphere dominates, as is the case for all space-geodetic observing techniques including VLBI, a convex shape as in Fig. 13.8 prevails. For radiation to be observed on Earth, this phenomenon is of course valid in the same way.

In principle, the relationship of Eq. 13.47 is valid only for a monochromatic wave and thus for the propagation of the phase, the so-called phase velocity v_{ph}. The propagation of a group of waves resulting from a superposition of multiple frequencies is affected by a partial retardation of the phases depending on the wavelength λ originating from *Rayleigh* scattering. This leads to a group velocity v_g with

$$v_g = v_{ph} - \lambda \frac{dv_{ph}}{d\lambda} \tag{13.48}$$

According to Eqs. 13.47 and 13.48 the group refraction index n_g can then be given depending on the frequency ν as

$$n_g = n_{ph} + \nu \frac{dn_{ph}}{d\nu} \tag{13.49}$$

(e.g., Marini and Murray, 1973).

13.5.1. Neutral atmosphere

The refractive effects of the neutral atmosphere are caused by induced and permanent dipole moments of water vapor as well as induced dipole moments of the non-water vapor elements (Owens, 1967). The refraction of the neutral atmosphere in the radio frequency domain is non-dispersive. For this reason, the additive term in Eq. 13.49 vanishes and

$$n_g = n_{ph} \tag{13.50}$$

As written in Eq. 13.46, the electrical path length L is the integral over the refractive indices along the geometric path length S. Since we are interested in the extra path length ΔL exceeding the straight line in vacuum G, we can write

$$\Delta L = \int_S n_g \, ds - G. \tag{13.51}$$

Because n_g is always a little bit above 1, around 1.0003 to be more exact, we can use the refraction number N with

$$N = (n_g - 1) \cdot 10^6 \tag{13.52}$$

and perform the transformations

$$\Delta L = \int_S n_g \, ds - \int_S 1 \, ds + S - G = \int_S (n_g - 1) \, ds + S - G = 10^{-6} \int_S N \, ds + (S - G) \tag{13.53}$$

$S-G$ is the purely geometric detour through the bending. Obviously $S-G=0$ for observations at zenith. For the assumption of stratified atmospheric layers, $S-G$ is approximately 10 cm for observations at 5° elevation. In most refraction models, the $S-G$ contribution is considered in the transformations from zenith to arbitrary elevation angles, i.e., in the mapping functions (below).

In terms of an extra time delay $\Delta\tau^{atm}$ and an extra total phase contribution $\Delta\Phi^{atm}$, Eq. 13.53 converts to

$$\Delta\tau^{atm} = \frac{10^{-6}}{c} \int_S N \, ds + \frac{(S - G)}{c} \tag{13.54}$$

and

$$\Delta\Phi^{atm} = \frac{2\pi}{\lambda}\, 10^{-6} \int_S N\, ds + \frac{2\pi}{\lambda}(S - G) \tag{13.55}$$

respectively.

There have been a few derivations of N, e.g., by Smith and Weintraub (1953) or Thayer (1974), but for VLBI the fundamental problem persists that these numbers need to be determined along the geometrical path to solve the integral in the above equations, which is close to impossible for the required level of accuracy. For this reason, another approach is chosen. A common way of modeling refraction in the neutral atmosphere is the separation of $\Delta\tau^{atm}$ into a hydrostatic and a water vapor part. The latter is often called the wet part. Both are separated further into a component reflecting the refraction effect in zenith direction and a mapping function for the conversion into the direction of the observation. The delay contribution $\Delta\tau^{atm}(\varepsilon)$ is commonly represented as extra path length $\Delta L^{atm}(\varepsilon)$ which is easily transformable to a time unit quantity through the speed of light. If the model only considers a dependency on elevation angle ε, it can be composed as

$$\Delta L^{atm}(\varepsilon) = \Delta L_h^z \cdot m_h(\varepsilon) + \Delta L_w^z \cdot m_w(\varepsilon). \tag{13.56}$$

$m_h(\varepsilon)$ and $m_w(\varepsilon)$ are the mapping functions which are applied to map the extra path lengths or their time equivalents in zenith direction onto any non-zenith elevation angle ε. A simple $1/\sin\varepsilon$ scaling is not sufficient, in particular at lower elevations because the curvature of the atmosphere as a consequence of the curvature of the Earth's surface would not be taken into account. Most modern mapping functions thus are a function of ε and of a finite continuous fractional form with coefficients (a, b, c) (Herring, 1992). They are specific for the hydrostatic (Eq. 13.57) and the wet component (Eq. 13.58) because the thickness of the effective atmosphere is more than 10 km in the first case and only about 2 km in the latter (Niell, 2000).

$$m_h(\varepsilon) = \frac{1 + \dfrac{a_h}{1 + \dfrac{b_h}{1 + c_h}}}{\sin\varepsilon + \dfrac{a_h}{\sin\varepsilon + \dfrac{b_h}{\sin\varepsilon + c_h}}} \tag{13.57}$$

$$m_w(\varepsilon) = \frac{1 + \dfrac{a_w}{1 + \dfrac{b_w}{1 + c_w}}}{\sin\varepsilon + \dfrac{a_w}{\sin\varepsilon + \dfrac{b_w}{\sin\varepsilon + c_w}}} \tag{13.58}$$

While in earlier mapping functions the coefficients mostly depended only on the geographic latitude of the telescope (Herring, 1992; Niell, 1996), modern mapping functions are inferred from numerical weather models (J. Böhm and Schuh, 2004; J. Böhm et al., 2006; Landskron and J. Böhm, 2018).

The importance of the mapping functions becomes clear when we consider that at lower elevation angles, the signals pass through multiples of the atmosphere thickness in zenith direction and that they also multiply any error in the zenith delays. It is thus convenient to relate the elevation angles, at least roughly, with the number of atmospheres to be passed (e.g., Dodson et al., 1996). A few of these numbers are listed in Tab. 13.2.

Table 13.2: Relationships of elevation angles and number of atmospheres to be passed.

Elevation angle	Number of atmospheres
90°	1
30°	2
20°	3
10°	5,7
5°	11,5
3°	19,1

The hydrostatic component of the extra path length in zenith direction ΔL_h^z follows a development of (Saastamoinen, 1972; Saastamoinen, 1973a; Saastamoinen, 1973b), finally formulated by Davis et al. (1985) as

$$\Delta L_h^z = 0.0022768\,[\mathrm{m\,hPa^{-1}}]\,\frac{P_\circ}{(1 - 0.00266\,\cos 2\phi\, - 0.00028\,H)}. \tag{13.59}$$

It predominantly depends on the barometric pressure at the height of the VLBI reference point P_0 [in hPa] and to a very small degree on latitude ϕ and the height above the geoid H. At sea level this corresponds to roughly 2.28 m extra path length or 7.6 ns.

The hydrostatic path delay for each observation can be applied a priori according to

$$\Delta\tau_{atm^h}(\varepsilon) = \frac{1}{c}\cdot \Delta L_h^z \cdot m_h(\varepsilon). \tag{13.60}$$

A determination of the wet component along the line of sight ΔL_w with sufficient accuracy regularly fails because it would need a solution for an integral along the path length over the partial pressure of the water vapor e, the temperature T and some compressibility factor $Z_w \approx 1$ (Owens, 1967) scaled by the mean temperature T_m according to the model developed by Elgered (1982):

$$\Delta L_w = [1 + 6\times 10^{-5}\,T_m]\,3.754\int_S \frac{e}{T^2}\,Z_w^{-1}\,ds. \tag{13.61}$$

The reason is that the water vapor in the atmosphere is highly variable in space and time.

Our inability to do a forward modeling of the wet component with sufficient accuracy leads to the generally applied strategy to just estimate an extra delay in zenith direction attributed to water vapor $\Delta\tau_w^z = \Delta L_w^z/c$ after the observations have been calibrated with a model hydrostatic contribution according to Eq. 13.60. More details on the background of refraction in the neutral atmosphere in VLBI can be found in J. Böhm et al. (2013).

13.5.2. Charged atmosphere

In the radio frequency domain, electrically charged particles have severe effects on the radio sources' radiation from their origin to their detection by radio telescopes such as retardation and rotation of the polarization plane (See App. C). Since these effects are frequency dependent the charged particles are said to cause dispersive effects. With our radio telescopes, we encounter dispersion as the integral of all charged particles along the electrical path in the direct vicinity of the source, the interstellar plasma, the solar corona (if rays are passing close enough), and finally of the Earth's charged atmosphere, which is commonly called the ionosphere.

Except of special tests of solar corona effects (Soja et al., 2014), geodetic VLBI has to deal mainly with the Earth's charged atmosphere. In this context, Eq. 13.49 gave the group refraction index n_g as the sum of the monochromatic phase velocity n_{np} and the frequency-dependent changes. In terms of the observing frequency ν, the first summand depends on the electron content EC with

$$n_{ph}^{ion} = 1 - \left(\frac{40.3}{\nu^2} + \frac{74.1 \times 10^6}{\nu^3}\right) \cdot EC \tag{13.62}$$

(Alizadeh et al., 2013; Brunner and Gu, 1991). In most applications, the second term can be ignored. The total electrical path length L_{ph}^{ion} is again the integral over the geometric path S

$$L^{ion} = \int_S n_{ph}^{ion}\, ds = \int_S \left[1 - \left(\frac{40,3}{\nu^2} + \frac{74,1 \times 10^6}{\nu^3}\right) \cdot EC\right] ds \tag{13.63}$$

which can be reformulated as

$$L_{ph}^{ion} = S - \left(\frac{40,3}{\nu^2} + \frac{74,1 \times 10^6}{\nu^3}\right) \int_S EC\, ds. \tag{13.64}$$

With the integral now purely applicable to the electron content and resulting in the total electron content along the geometrical path length TEC, the extra path length ΔL_{ph}^{ion} is

$$\Delta L_{ph}^{ion} = -\left(\frac{40,3}{\nu^2} + \frac{74,1 \times 10^6}{\nu^3}\right) \cdot TEC + (S - G) \tag{13.65}$$

and the correction to the phase delay $\Delta\tau_{ph}^{ion}$ and to the total phase is $\Delta\Phi^{ion}$ are

$$\Delta\tau_{ph}^{ion} = -\left(\frac{40,3}{c\cdot\nu^2} + \frac{74,1\times10^6}{c\cdot\nu^3}\right)\cdot TEC + \frac{(S-G)}{c} \tag{13.66}$$

and

$$\Delta\Phi^{ion} = -\frac{2\pi}{\lambda}\left(\frac{40,3}{\nu^2} + \frac{74,1\times10^6}{\nu^3}\right)\cdot TEC + \frac{2\pi}{\lambda}(S-G), \tag{13.67}$$

respectively. If we replace the wavelength λ by f/ν we find that

$$\Delta\Phi^{ion} = -\frac{2\pi}{c}\left(\frac{40,3}{\nu} + \frac{74,1\times10^6}{\nu^2}\right)\cdot TEC + \frac{2\pi\nu}{c}(S-G), \tag{13.68}$$

which shows us that the phase acceleration is inverse to the frequency ν.

The refraction index for a group of waves n_{gr}^{ion} looks quite similar to that of the monochromatic phase:

$$n_{gr}^{ion} = 1 + \left(\frac{40.3}{\nu^2} + 2\cdot\frac{74.1\times10^6}{\nu^3}\right)\cdot EC \tag{13.69}$$

which leads to an extra delay $\Delta\tau_{gr}^{ion}$

$$\Delta\tau_{gr}^{ion} = \left(\frac{40.3}{c\cdot\nu^2} + 2\cdot\frac{74.1\times10^6}{c\cdot\nu^3}\right)\cdot TEC + \frac{(S-G)}{c} \tag{13.70}$$

However, the signs of Eq. 13.68 and Eq. 13.70 are different, minus for the phase (in fact the phase is accelerated through the ionosphere) and plus for the time delay. For this reason, the ionosphere correction for the phase delay and total phase need to be added instead of subtracted as for the group delay. The retardation of the wavefront is inversely proportional to the square of the frequency ν.

The electron content is measured in electrons per square meter. The total electron content is thus the number of electrons in a column with a base of 1 m^2. The electron content varies predominantly through the sun's radiation and its variability. At night the TEC in zenith direction is approx. 5×10^{16} electrons, while at daytime it is approx. 5×10^{17}. 1 TECU (TEC Unit) is 1×10^{16} electrons. At 8.4 GHz (X band) the extra group delay (for a single location) is approx. 1 nanosecond (ns)(0.3 m) at daytime and 100 picoseconds (ps)(3 cm) at nighttime, while at 2.3 GHz (S band) it is 13 ns and 1.3 ns, respectively.

As for the integrals of the neutral atmosphere, the difficulty with the charged atmosphere is that we need the total electron content along the geometric path length with sufficient accuracy. However, since the dispersive effect can be modelled straight forwardly, VLBI observations have always been carried out with two separate frequencies (at X band and S band, Sec. 3) for calibration of the primary observing frequency, i.e., X band. The correction

$\Delta\tau^X_{ion}$ for the X band observable is easily computed (Alizadeh et al., 2013) by

$$\Delta\tau^X_{ion} = (\tau_S - \tau_X) \cdot \frac{\nu_S^2}{\nu_S^2 - \nu_X^2} \tag{13.71}$$

where τ_S and τ_X are the measured multiband group delays and ν_S and ν_X the observed frequencies at S band and X band, respectively. N.B.: To compute the ionospheric contribution, which by the way is the quantity stored in vgosDB files, the sign has to be toggled.

Since the observed frequency of each band actually results from a combination of several frequencies through bandwidth synthesis, these frequencies need to be computed as effective frequencies which are the weighted mean over all channels/sub-bands N

$$\nu_{gr} = \sqrt{\frac{\sum\limits_{i=1}^{N} \rho_i \cdot \sum\limits_{i=1}^{N} \rho_i(\nu_i - \nu_0)^2 - \left(\sum\limits_{i=1}^{N} \rho_i(\nu_i - \nu_0)\right)^2}{\sum\limits_{i=1}^{N} \rho_i(\nu_i - \nu_0) \cdot \sum\limits_{i=1}^{N} \frac{\rho_i}{\nu_i} - \sum\limits_{i=1}^{N} \rho_i \cdot \sum\limits_{i=1}^{N} \rho_i \frac{\nu_i - \nu_0}{\nu_i}}} \tag{13.72}$$

ρ_i are the weights of individual channels/sub-bands. These can be computed from the number of correlated bits, from SNR, or from the fringe amplitude which is readily available in standard vgosDB files. ν_0 is the reference frequency of the bandwidth synthesis setup, e.g., channel/sub-band #1 (Herring, 1983; Petrov, 2001).

The frequencies ν_i themselves need a closer look. Regularly, a frequency sub-band/channel is named by its nominal frequency together with its sideband, which may be lower and upper sideband (LSB/USB). Each S/X sub-band has a bandwidth of 4, 8, or 16 MHz. The effective frequency of a single channel then lies in the middle of this bandwidth. For this reason the computation of the channel frequency needs to add half of the channel bandwidth to the nominal frequency for USB channels and to subtract half for LSB channels. If the channel bandwidth is not readily available, it can also be computed by (sample rate)/4.

A complication now arises from the fact that at X band the two outer sub-bands of the nominal eight sub-bands are in fact made up of a lower and an upper sideband (LSB/USB) channel doubling the effective border sub-bands resulting in a total of ten channels for eight sub-bands. In the fringe fitting process the data of the outer channels are concatenated. In this case the effective sub-band frequency is the nominal channel frequency. No adding or subtracting half the channel bandwidth is required.

As a result of the concatenation of LSB/USB channels, the fringe amplitude does not increase but the SNR of the sub-bands does (through the extended number of samples). If everything was normal during observations, channel amplitudes are more or less constant for all channels including those of the LSB channels. This is the reason why *fourfit* fringe plots show similar amplitudes for the eight sub-bands reported.

The computation of the effective frequencies has to take care that channels/sub-bands, which have not produced proper fringes or which have an inappropriate number of correlated samples, are treated properly, i.e., are excluded. In vgosDB files for X band, we find „ChanAmpPhase“ with the dimension (2 × 8) where the first line contains the amplitude [0, 1] and the second line the phase [-180, 180]. Through this, vgosDB files contain the amplitudes of all sub-bands with a structure based on reference frequencies. In addition, the entry "Observables.ChannelInfo_bX.NumSamples" ('# of samples by sideband and channel') also has the dimension (2 × 8). However, here the first line contains the LSB number of samples and the second line the USB number of samples. Since we only have two sub-bands with LSB data, only elements (1,1) and (1,8) contain non-zero values in the first line while all others are zero. The second line should be fully populated with non-zeros unless we encountered a channel drop-out.

In the latest stage of developments, the weight of each sub-band i is then computed from

$$\rho_i = (N_{i-LSB} + N_{i-USB}) * Amp_i \tag{13.73}$$

where N_{i-LSB} and N_{i-USB} are the number of samples in LSB and USB, respectively. Amp_i is the amplitude of each of the sub-bands.

To determine the formal error of $\Delta\tau^X_{ion}$, we first form the partial derivatives with respect to the parameters of Eq. 13.71 which are

$$\frac{d\Delta\tau^X_{ion}}{d\tau_S} = \frac{\nu_S^2}{\nu_S^2 - \nu_X^2} \tag{13.74}$$

$$\frac{d\Delta\tau^X_{ion}}{d\tau_X} = -\frac{\nu_S^2}{\nu_S^2 - \nu_X^2} \tag{13.75}$$

$$\frac{d\Delta\tau^X_{ion}}{d\nu_X} = (\tau_S - \tau_X) \cdot [2\nu_S \cdot \frac{1}{\nu_S^2 - \nu_X^2} - \nu_S^2 \cdot \frac{1}{(\nu_S^2 - \nu_X^2)^2} \cdot 2\nu_S] \tag{13.76}$$

and

$$\frac{d\Delta\tau^X_{ion}}{d\nu_S} = (\tau_S - \tau_X) \cdot [\nu_S^2 \cdot \frac{1}{(\nu_S^2 - \nu_X^2)^2} \cdot 2\nu_X]. \tag{13.77}$$

With simple uncorrelated error propagation, the formal error of the ionospheric calibration then is

$$\sigma_{\tau^X_{ion}} = \sqrt{\left(\frac{d\Delta\tau^X_{ion}}{d\tau_S}\right)^2 \cdot \sigma^2_{\tau_S} + \left(\frac{d\Delta\tau^X_{ion}}{d\tau_X}\right)^2 \cdot \sigma^2_{\tau_X} + \left(\frac{d\Delta\tau^X_{ion}}{d\nu_S}\right)^2 \cdot \sigma^2_{\nu_S} + \left(\frac{d\Delta\tau^X_{ion}}{d\nu_X}\right)^2 \cdot \sigma^2_{\nu_X}}. \tag{13.78}$$

Now, the question may arise, how sensitive the ionospheric correction reacts to errors in the effective frequencies. For this purpose, we take the derivatives with respect to ν as in Eqs. 13.76 and 13.77 and multiply them with the respective shift $\Delta\nu$ yielding

$$\Delta \tau_{ion}^{X} = \frac{d\tau_{ion}^{X}}{d\nu} \cdot \Delta\nu. \tag{13.79}$$

A numerical example can be constructed by asking how accurate $\Delta\nu$ has to be determined for a 1 ps sensitivity limit. For this, we assume an ionospheric delay contribution of 13 ns at S band and 1 ns at X band (see example above) and an observation at 10° elevation scaling the delay difference term in the partial derivatives to 69.1 ns. With the partial derivatives of 5.5×10^{-11} and 2.1×10^{-11} with respect to ν_S and ν_X, respectively, we find that the effective frequency has to be determined better than 0.26 MHz for S band and 0.69 MHz for X band. The requirements are even more stringent for observations at even lower elevation angles but relax by a factor of up to five for observations near zenith (at both telescopes).

More details on the nature of ionospheric refraction are described in Alizadeh et al. (2013).

13.6. Further model contributions

For the final geometric time delay in the terrestrial system, all known contributions have to be added to Eq. 13.45. The first ones are those of refraction through the ionosphere $\Delta\tau_{ion}$ according to Eq. 13.71 and through the neutral atmosphere $\Delta\tau_{atm^h}$ according to Eq. 13.60. At this point we also have to take into account that besides the aberration effects in vacuum embedded in Eq. 13.45, the path through the charged and neutral atmosphere at telescope 1 also permits telescope 2 to change its position due to the rotation of the Earth and its movement in space before the signal arrives there causing an additional aberration contribution $\Delta\tau_{abb^{atm}}$

$$\Delta\tau_{abb^{atm}} = \Delta\tau_{atm_{(1)}} \cdot \frac{\mathbf{K} \cdot (\mathbf{w}_{(2)} - \mathbf{w}_{(i)})}{c}. \tag{13.80}$$

Further known corrections to be applied to the theoretical delay are tidal displacements and loading effects caused by the variability in the load forces of the oceans and the atmosphere. A complete list of them can be found in the Conventions of the International Earth Rotation and Reference Systems Service (IERS) (Petit and Luzum, 2010) with many references to the original scientific publications.

These effects all lead to displacements of the radio telescopes $\Delta\mathbf{X}_{mod}$ on the inter-annual to sub-daily time scale. Since they are all in the linear domain for the analysis at hand, their impact can be easily transferred to the time delay domain through computing

$$\Delta\tau_{mod} = \frac{\partial\tau}{\partial\mathbf{X}_{mod}} \cdot \Delta\mathbf{X}_{mod}. \tag{13.81}$$

It should be noted that the latter contributions, which are coordinate displacements in their original form, do not necessarily need to be applied at this stage. As adopted in some software packages they may also be applied to the a priori coordinates before going through the full relativistic formulation in Eq. 13.45 (Fig. 13.9).

The formulation of the final theoretical or a priori delay depends on whether the corrections mentioned above are actually applied to the observed delay, to the geometry or to the a priori delay. In

$$\tau_{apriori} = \tau_{vac} + \Delta\tau_{abb^{atm}} + \Delta\tau_{E.T.} + \Delta\tau_{P.T.} + \Delta\tau_{O.L.} + \Delta\tau_{A.L.} \tag{13.82}$$

the contributions $\Delta\tau_{E.T.}$ of the Earth tides, $\Delta\tau_{P.T.}$ of the pole tide, $\Delta\tau_{O.L.}$ of ocean loading, and $\Delta\tau_{A.L.}$ of atmospheric loading are applied to the theoretical delay. In contrast to this, $\Delta\tau_{T.E.}$ the contribution of the telescopes' thermal expansion, $\Delta\tau_{G.D.}$ of the telescopes' gravitational deformations, and $\Delta\tau_{Ins}$ of the telescopes' cable delays (Sec. 7.1) are applied as corrections to the observed delay

$$\tau_{obs^{corr}} = \tau_{obs} - \Delta\tau_{ion} - \Delta\tau_{atm^h} - \Delta\tau_{T.E.} - \Delta\tau_{G.D.} - \Delta\tau_{Cable} \tag{13.83}$$

All the $\Delta\tau_*$ have of course to be computed as differences related to telescope 1 and 2. Applying the respective effects to the theoretical delays provides the **c** (computed $\equiv \tau_{apriori}$) vector and the vector of observations **o** (observed $\equiv \tau_{obs^{corr}}$). Forming the differences produces the linearized (abbreviated) observation vector

$$\mathbf{y} = \mathbf{o} - \mathbf{c} \tag{13.84}$$

for the parameter estimation.

13.7. Parameter estimation

Parameter estimation is generally carried out with the abbreviated observation vector (Eq. 13.84). This implies according to the τ_{vac} in Eq. 13.82 that a priori telescope coordinates, Earth orientation parameters and radio source positions have also been "subtracted" already. This means that we will estimate only adjustments to these a priori values. Consequently, the adjustment is then just carried out in the linear regime where the fit and the estimation are reduced to small magnitudes. At the end of the process, the estimates are added again to the a prioris producing the total values of the parameters.

Various estimation schemes exist to determine the parameters of interest from VLBI multi-band group delay observations. The primary difference is how these handle the stochastic nature of the clocks and in particular the atmosphere. Kalman Filter (Herring et al., 1990; Kalman, 1960; Nilsson et al., 2015; Soja et al., 2015), least-squares collocation (Titov, 2000), square-root information filters (Bolotin, 2000) treat the wet part of the atmosphere and also the clock behavior as a purely stochastic process. However, most common are standard least-squares adjustments in Gauß-Markov models. It should be mentioned here, that in comparison campaigns it could not be substantiated that any of these estimation processes produces superior results. A drawback of the Filter solutions is, that combination of the results on the basis of normal equation systems is not possible directly, and for least-squares collocation, the extraction of a partly resolved normal equation system is quite complicated.

To structure the introductory description of the least-squares adjustment in a Gauß-Markov model, we first look at the stochastic model of the observations. This determines the weight which each observation contributes to the least-squares adjustment (Sec. 13.7.1). Thereafter, we run the Gauß-Markov model (Sec. 13.7.2) with the (abbreviated) observations. The theory behind this is described, for example, by Koch (1999).

13.7.1. Stochastic model

The anchor point of the VLBI stochastic model is the aggregate of the standard deviations or their squares, the variances, of the observations, σ_i or σ_i^2, respectively. In general, the observations cannot be considered as independent but are correlated in a statistical sense. This fact can be described by correlation coefficients $\rho_{i,j}$ ($\rho = [-1,1]$). With respect to the sequence of observations y_i, these quantities are ordered in the covariance matrix of the observations $\mathbf{\Sigma_{yy}}$ for n observations with the dimension $n \times n$

$$\mathbf{\Sigma_{yy}} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \dots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \dots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \rho_{n2}\sigma_n\sigma_2 & \dots & \sigma_n^2 \end{pmatrix} \tag{13.85}$$

The elements $\rho_{ij}\sigma_i\sigma_j$ are the covariances σ_{ij}. With variances and covariances available, we can determine the statistical correlation coefficients ρ_{ij} by

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \cdot \sigma_j}. \tag{13.86}$$

However, this is normally mostly only applied to the elements of the covariance matrix of the parameters a posteriori, i.e., after inversion of the normal equation matrix (Eq. 13.106).

It should be noted that often no absolute precision information of the observables is available and only relative quantities can be employed. This, for example, is the case with VLBI group and phase delays as well as their rates which are derived from Signal-to-noise considerations initially (Sec. 3). It is therefore appropriate to decompose the covariance matrix into an a priori variance factor σ_0 and a cofactor matrix a priori $\mathbf{Q_{yy}}$ with

$$\mathbf{\Sigma_{yy}} = \sigma_0^2\,\mathbf{Q_{yy}}. \tag{13.87}$$

Although many studies exist on the correlations between observables, they are still neglected in operational VLBI analyses. Various attempts have been made to cope with this problem, also for other space geodetic techniques, but solutions tended to be computationally expensive and thus not suitable for operational application (Romero-Wolf et al., 2012). Some time ago, a very promising approach was published by using turbulence theory as the driver for the correlations. Due to its extremely good results and even more so that the algorithm has favorable runtime costs, it is well suited for operational applications (Halsig et al., 2016).

Resulting from the neglect of the correlations, the co-factor matrix of the observations $\mathbf{Q_{yy}}$ and, thus, the covariance matrix $\mathbf{\Sigma_{yy}}$ regularly only contain diagonal elements, i.e., the variances of the observations $q_{ii} = \sigma_i^2$ which are deduced from the signal-to-noise ratio of the correlation process (Sec. 10.3).

$$\mathbf{\Sigma_{yy}} = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix} \tag{13.88}$$

Writing down Eq. 13.87 for the uncorrelated case, we find

$$\mathbf{\Sigma_{yy}} = \sigma_0^2\,\mathbf{Q_{yy}} = \sigma_0^2 \begin{pmatrix} \sigma_1^2/\sigma_0^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2/\sigma_0^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2/\sigma_0^2 \end{pmatrix} \tag{13.89}$$

From this we can now move to the weight matrix $\mathbf{P}$ which is the inverse of the cofactor matrix

$$\mathbf{P} = \mathbf{Q_{yy}^{-1}} = \begin{pmatrix} \sigma_0^2/\sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_0^2/\sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_0^2/\sigma_n^2 \end{pmatrix} = \begin{pmatrix} p_{11} & 0 & \dots & 0 \\ 0 & p_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p_{nn} \end{pmatrix}. \tag{13.90}$$

In the literature, there is an undecided discussion whether the elements of the $\mathbf{P}$ matrix, the p_{ii}, have a dimension or are dimensionless. The question arises from the dimension of the variance of unit weight a posteriori (Eq. 13.108. If the weights are dimensionless, which would be the intuitive answer because of the division of two similar quantities, the variance of unit weight a posteriori has the dimension of the observations squared and should always be stated explicitly. However, some colleagues argue that the σ_0^2 is dimensionless, among other arguments that it can also be set to 1. In this case, the quadratic term (Eq. 13.109) and thus the variance of unit weight becomes dimensionless which is the predominant view in the statistics community. In any case, the numbers themselves do not change but it should be emphasized that this is one of the reasons why great care has to be taken to keep the dimensions in the analysis programs consistent from the beginning to the end.

Finally, we should briefly look at the global test of the χ^2 factor which is the quotient of the variance factor a posteriori $\tilde{\sigma}_0^2$ (Eq. 13.108) and the variance factor a priori σ_0^2

$$\chi^2 = \frac{\tilde{\sigma}_0^2}{\sigma_0^2}. \tag{13.91}$$

This should be unity or at least close to it proving that the variance of unit weight of the observations a priori matches that of the least squares fit. If you run a VLBI solution just with the standard deviations from the correlation and fringe fitting process, χ^2 is seldom near the required unity. This means that the latter are too optimistic. For this reason, the most common way of handling this problem in operational analyses, i.e., for standard solutions of the IVS, is by inflating the input variances to the least squares adjustment with an additive constant σ^2_{add} (additive noise) (Eq. 13.93). For legacy S/X observations, the input variances refer to the observations corrected for ionospheric dispersion effects. These are computed with

$$\sigma^2_{i,corr} = \sigma^2_{i,X} + \sigma^2_{i,ion} \tag{13.92}$$

where $\sigma^2_{i,X}$ are the SNR derived variances of the X band observations and $\sigma^2_{i,ion}$ are the variances of the ionospheric calibrations which are derived by variance propagation of the X band and S band variances applied to Eq. 13.71. The inflated variances $\sigma^2_{i,infl}$ are then

$$\sigma^2_{i,infl} = \sigma^2_{i,corr} + \sigma^2_{add} \tag{13.93}$$

to achieve that the aggregate inflated input variances to the adjustment $\sigma^2_{0,infl}$ match those a posteriori and produce the χ^2 in Eq. 13.91 to become close to 1. For practical reasons, it is convenient to set the variance factor a priori $\sigma_0^2 = 1$. The additive noise can, for example, be computed baseline-wise or for the entire session.

A further refinement of this approach is the introduction of elevation dependent noise (Gipson et al., 2008). Here, noise contributions for each telescope are added as a function of the respective elevation angle ε with

$$\sigma^2_{i,infl} = \sigma^2_{i,corr} + + \left(\frac{c_i}{\sin(\varepsilon_i)}\right)^2 + \left(\frac{c_j}{\sin(\varepsilon_j)}\right)^2 [+\sigma^2_{add}] \tag{13.94}$$

with, for example, constants $c_{i,j} = 6$ ps for each telescope. Even here, a session-dependent constant additive noise could be added if necessary.

Although the application of additive noise solves the imbalance of the variances numerically, the deficit of the approach is twofold. The first is that the error contributions are computed to make the χ^2 in Eq. 13.91 unity, thus they are purely empirical and lack any physical explanation. The second deficit is that correlations between observations are still ignored entirely.

13.7.2. Gauß-Markov model

The estimation process in a Gauß-Markov model can be followed easily in Fig. 13.9, where the raw observations produced in the correlation and fringe fitting process are first corrected for all known effects mentioned before including the hydrostatic part of the atmosphere. On the right hand side, the radio source positions are transformed into position of date by the precession/nutation rotations composed in the rotation matrix $\mathbf{Q}(t)$ in Eq. 13.114. The a priori telescope coordinates on the right hand side are mostly those of the ITRF (Sec. 13.1.1) which are transformed to the epoch of the observations by applying the corrections for the station velocity components in the ITRF and the rotations from the catalog frame to the frame of date through the Earth rotation parameters, polar motion and the Earth's phase of rotation. Further corrections to the coordinates are the effects of the Earth tides, the pole tide, oceanic and atmospheric loading effects as mentioned above.

The classical least squares adjustment in a Gauß-Markov model reads (Koch, 1999)

$$\mathbf{y} = \mathbf{A} \cdot \mathbf{x} + \mathbf{v} \tag{13.95}$$

where $\mathbf{y}$ is the n × 1 vector of the abbreviated observations (Eq. 13.84), $\mathbf{x}$ is the m × 1 vector of the unknown parameters and $\mathbf{v}$ is the n × 1 vector of the residuals (Koch, 1999). Abbreviated observations, in the VLBI case, are the original phase, delay, or delay rate observations minus the respective values computed from a priori models, also called "observed minus computed" (O-C) (Fig. 13.9). $\mathbf{A}$ is the n × m Jacobian or linearized design matrix containing the partial derivatives of the observations w.r.t. the parameters to be estimated

$$\mathbf{A}(i,j) = \frac{\partial \tau_{obs,i}}{\partial x_j}. \tag{13.96}$$

Translating this in practical terms, we start with an equation (for each observation), which relates observations and parameters, called observation equation. The VLBI observation equation is based on the initial formulation in Eq. 13.29 with the following extensions. Since the two clocks are never synchronized to the accuracy needed, the geometric model is expanded for the respective, time dependent clock offset. The clock behavior is time dependent and we can formulate the clock contributions $\Delta\tau_{clo}$ as second order polynomials

$$\begin{aligned} \Delta\tau(t)_{clo} = \; & -(\mathcal{T}_0^{(1)} + \mathcal{T}_1^{(1)} \cdot (t-t_0) + \mathcal{T}_2^{(1)} \cdot (t-t_0)^2) \\ & +(\mathcal{T}_0^{(2)} + \mathcal{T}_1^{(2)} \cdot (t-t_0) + \mathcal{T}_2^{(2)} \cdot (t-t_0)^2) \end{aligned} \tag{13.97}$$

to be added to Eq. 13.29. $\mathcal{T}_j^*$ are the parameters of the clock polynomial at telescope (1) or (2).

The contribution of the refraction caused by water vapor, which we like to estimate, is expressed as the product of a zenith effect and a mapping function (Sec. 13.5). In a simple

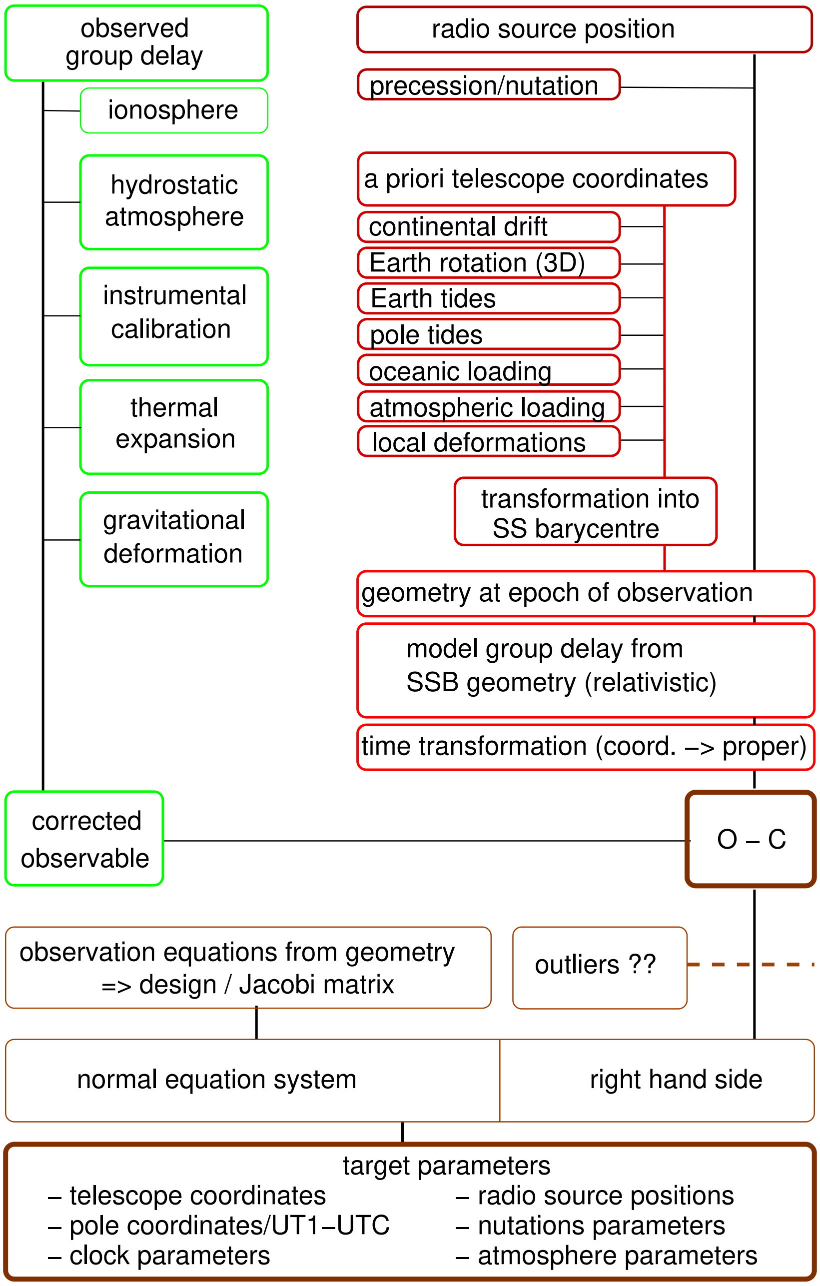


Figure 13.9: Standard Gauß-Markov model VLBI analysis data flow.

form, we can then write the observation equation as

$$
\begin{aligned}
\tau_{obs}(t) = \quad & -\tfrac{1}{c}\,\mathbf{b}\cdot\mathbf{W}(x_p,y_p)\cdot\mathbf{S}(\theta)\cdot\mathbf{Q}(X,Y)\cdot\mathbf{k} \\
& -(\mathscr{T}_0^{(1)}+\mathscr{T}_1^{(1)}\cdot(t-t_0)+\mathscr{T}_2^{(1)}\cdot(t-t_0)^2) \\
& +(\mathscr{T}_0^{(2)}+\mathscr{T}_1^{(2)}\cdot(t-t_0)+\mathscr{T}_2^{2}\cdot(t-t_0)^2) \\
& +m_w^{(1)}\cdot\mathscr{AT}^1+ \\
& +m_w^{(2)}\cdot\mathscr{AT}^2+ \\
& +\tau_{corr}+v
\end{aligned}
\tag{13.98}
$$

with

$\mathbf{k}$	unit vector in source direction (Eq. 13.31) its three components k_i
$\mathbf{b}$	baseline vector (Eq. 13.30)
c	velocity of light
$\mathbf{W}$	rotation matrix for polar motion
	with the angular arguments x_p and y_p
$\mathbf{S}$	rotation matrix for the daily spin of the Earth
	with the angular argument θ/UT1
$\mathbf{Q}$	transformation matrix for precession and nutation
	with the angular arguments X and Y
$\mathscr{T}_j^*$	parameters of the clock polynomial at telescope 1 or 2
m_w^*	mapping function of the wet part of the atmosphere at station 1 or 2
$\mathscr{A}\mathscr{T}^*$	zenith wet delays at station (1) or (2)
τ_{corr}	sum of all corrections
v	residuals containing the measurement noise.

This equation is linearized by forming the observed minus computed (abbreviated) observations leaving only linear adjustments to the parameters of interest and forming the partial derivatives of the observations with respect to the parameters. This can then be decomposed to the fundamental Gauß-Markov model (Eq. 13.95) with $\mathbf{y} = \tau_{obs^{corr}} - \tau_{apriori}$.

The goal of the least-squares adjustment is to minimize the sum of squared residuals

$$\mathbf{v}^T\,\mathbf{P}\,\mathbf{v} \ldots \min \tag{13.99}$$

by finding the estimated parameters $\tilde{\mathbf{x}}$ which best match the abbreviated observation $\mathbf{y}$. This is achieved by formulating the normal equation system

$$\mathbf{N}\,\mathbf{x} = \mathbf{b} \tag{13.100}$$

where $\mathbf{N}$ is the *normal matrix* with

$$\mathbf{N} = (\mathbf{A}^T\,\boldsymbol{\Sigma}_{yy}^{-1}\mathbf{A}) \tag{13.101}$$

and $\mathbf{b}$ is the right hand side of the normal equation system with

$$\mathbf{b} = \mathbf{A}^T\,\boldsymbol{\Sigma}_{yy}^{-1}\,\mathbf{y} \tag{13.102}$$

Explicitly, the system reads

$$(\mathbf{A}^T\,\boldsymbol{\Sigma}_{yy}^{-1}\mathbf{A})\,\mathbf{x} = \mathbf{A}^T\,\boldsymbol{\Sigma}_{yy}^{-1}\,\mathbf{y} \tag{13.103}$$

$\boldsymbol{\Sigma}_{\mathbf{yy}}$ is the covariance matrix of the observations and is computed according to Eq. 13.85. The inverse of the covariance matrix is the weight matrix $\mathbf{P} = \boldsymbol{\Sigma}_{\mathbf{yy}}{}^{-1}$.

$\mathbf{N}$ needs to be regular to be invertible for achieving a solution (for background on datum definition see Sec. 13.8). The vector of the estimated parameters $\tilde{\mathbf{x}}$ is then computed with

$$\tilde{\mathbf{x}} = (\mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{y} \qquad (13.104)$$

The residuals $\mathbf{v}$ can be computed from

$$\mathbf{v} = \mathbf{A} \cdot \tilde{\mathbf{x}} - \mathbf{y}. \qquad (13.105)$$

The covariance matrix of the estimated parameters is

$$\mathbf{\Sigma}_{\tilde{x}\tilde{x}} = \tilde{\sigma}^2 (\mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{A})^{-1}. \qquad (13.106)$$

$$\mathbf{Q}_{\tilde{x}\tilde{x}} = (\mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{A})^{-1} \qquad (13.107)$$

is also called the cofactor matrix a posteriori. $\tilde{\sigma}^2$ is the variance of unit weight a posteriori and is computed as

$$\tilde{\sigma}_0^2 = \frac{\Omega}{n-u} = \frac{\mathbf{v}^T \mathbf{P} \mathbf{v}}{n-u} \qquad (13.108)$$

where n—u is the number of degrees of freedom (dof, number of observations minus number of unknowns in the least squares adjustment) and Ω results from

$$\Omega = \mathbf{v}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{v} \qquad (13.109)$$

which is also called the "weighted square sum of residuals". It has to be emphasized that n is actually the number of original observations plus the number of constraint equations introduced to stabilize the estimation of clock and atmosphere parameters.

It should be emphasized that in a least squares adjustment of VLBI observations, the clock parameters always have to be estimated because of the lack of synchronized clocks at far distant locations. However, since the clock parameters can only be estimated relative to each other, the parameters of a reference clock have to be fixed (conveniently to zero), i.e., eliminated from the parameter vector by deleting the respective columns from the design matrix. All other parameters of interest such as telescope coordinates, Earth orientation parameters, radio source positions, and zenith wet delays may be included in the vector of unknown parameters $\mathbf{x}$ more or less freely depending on the observation constellation. Nonetheless, certain datum considerations have to be taken into account when estimating telescope coordinates and radio source positions (Secs. 13.8 and 13.9).

Finally, it should be clarified that the residuals in Eq. 13.105 are rather corrections to the observations because they follow the initial Gauß-Markov model formulation (Eq. 13.95). To be able to interpret the residuals in the sense that a positive residual indicates an (abbreviated) observation which is too large, or vice versa for negative residuals, the $\mathbf{v}$ vector needs to multiplied by $-\mathbf{1}$ to produce genuine residuals.

13.7.3. Extension of the functional model

Over the years, it had been realized that the functional model of the least squares adjustments in Gauß-Markov models as formulated in Eq. 13.98 could be improved in several areas. The first one to mention is the increase of time resolution for the zenith wet delays and the clock parameters, but can also be applied to Earth rotation parameters as realized in the Vienna VLBI and Satellite Software, VieVS (Böhm et al., 2018). For this purpose, the parameter model is expanded with a number of additional parameters using continuous piece-wise linear functions (CPWLF) or linear splines (De Boor, 1978) of suitable time segments such as one hour or less for zenith wet delays and clocks, or 24 h at day breaks for ERP. In the case of the short time intervals of 1 h and below, the parameterization can be interpreted as a pseudo-stochastic modelling of moderately rapid changing phenomena.

CPWLFs come along as continuous piece-wise linear polygons (Fig. 13.10) where the polygons can be formulated in two different ways. The first one is an initial offset (functional value) at a reference epoch t_0 plus a new rate parameter for every segment of predefined duration (e.g., 20, 30 or 60 minutes). The functional values for each observation epoch t can be formulated according to Eq. 13.110 depending on the rates r_i and the segment limits t_i. This also serves as (part of) the VLBI observation equation which is the basis for the partial derivatives of the offset $f(t_0)$ and rate parameters r_i.

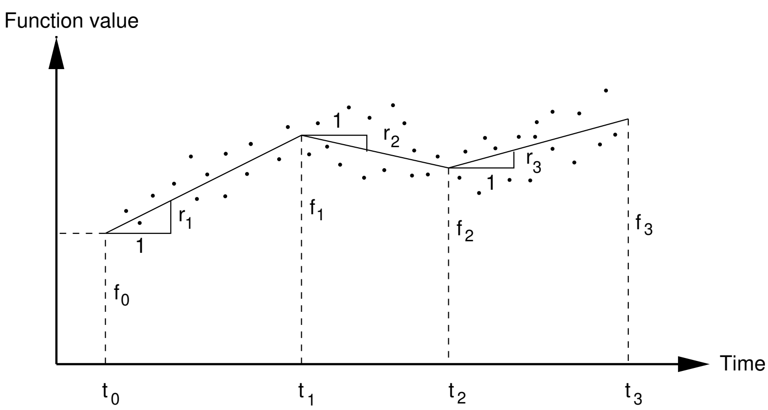


Figure 13.10: Continuous piece-wise linear offset parameter representation with observations (black dots) for a generic parameter.

$$f(t) \quad = \quad f(t_0) + r_1(t_1 - t_0) + r_2(t_2 - t_1) + ... + r_n(t - t_{n-1}) \qquad (13.110)$$

The functional values depend on the nature of the parameter such as $f(t) = \mathscr{T}$ and $f(t) = m_w^{(1)} \cdot \mathscr{AT}^1$ for the clock offsets and zenith wet delays, respectively, or the polar motion and dUT1 parts of Eq. 13.98. The rates also depend on the functional values f_i at the limits of the intervals.

$$r_i = \frac{f(t_i) - f(t_{i-1})}{t_i - t_{i-1}} \tag{13.111}$$

Consequently, we can introduce Eq. 13.111 in 13.110 and get a functional model purely based on the functional values (offsets) f_i (Eq. 13.112).

$$\begin{aligned} f(t) &= f(t_0) \\ &+ \frac{f(t_1) - f(t_0)}{t_1 - t_0}(t_1 - t_0) \\ &+ \frac{f(t_2) - f(t_1)}{t_2 - t_1}(t_2 - t_1) \\ &+ \frac{f(t_{i+1}) - f(t_i)}{t_{i+1} - t_i}(t - t_i) \\ &+ \ldots \end{aligned} \tag{13.112}$$

The approach using the functional values or offsets (Eq. 13.112) is equivalent to that applying and estimating the rates (Eq. 13.110). To expand the observation equation (Eq. 13.114) for these parameters, they are added in lines 4 and 5. The same mechanism also applies to the temporal resolution of the zenith wet delays for which the observation equation (Eq. 13.114) is expanded in lines 6 and 7.

For convenience, the reference epoch of the first relative clock offset parameter $f(t_0)$ has always been set to the start of the respective session in almost all VLBI analysis packages. This has the consequence that, due to error propagation, the formal errors of the estimated clock offsets increase with time as can be seen in Fig. 13.11 (bottom/red) part. Changing the clock reference epoch to the middle of the session improves the formal errors of the clock offset and rate parameters but has no effect on the other parameters of a VLBI solution. For more details on this issue see Nothnagel and Krásná (2021).

So far, the functional model of atmospheric refraction assumes isotropy and just contains the wet delay in zenith direction as one or more parameters to be estimated. However, this does not account for the anisotropy of refraction which may be caused, e.g., by the increased thickness of the atmospheric layers towards the equator. For this reason, MacMillan (1995) introduced gradients in the model

$$\begin{aligned} \tau_{grad}(t) = \; & +m_w^{(1)}(\varepsilon^{(1)}) \cdot \cot\varepsilon^{(1)} \cdot [\mathcal{G}_n^{(1)} \cos\alpha^{(1)} + \mathcal{G}_e^{(1)} \sin\alpha^{(1)}] \\ & +m_w^{(2)}(\varepsilon^{(2)}) \cdot \cot\varepsilon^{(2)} \cdot [\mathcal{G}_n^{(2)} \cos\alpha^{(2)} + \mathcal{G}_e^{(2)} \sin\alpha^{(2)}] \end{aligned} \tag{13.113}$$

It should be emphasized here, that any remaining unmodeled hydrostatic contribution

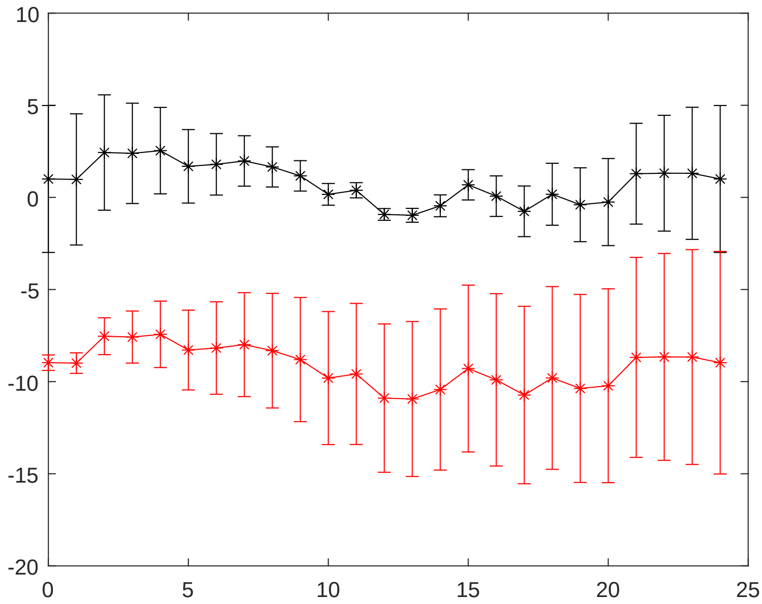


Figure 13.11: Estimates [cm] of clock piece-wise linear polygons for OV-VLBA in session 17NOV28XA. 1 cm $\hat{=}$ 33 ps. Units of x axis are hours since the session start. Bottom = Reference epoch at beginning of session, Top = reference epoch at middle of session. For more information see Nothnagel and Krásná (2021).

will be compensated for in the estimated wet component and in the gradients. The same applies to inaccurate pressure monitoring at the sites. For this reason the absolute values of the wet zenith delays may not be representative or even utterly wrong, e.g., if they are negative.

Taking all the augmentations of the functional model into account, the full observation equation in a Gauss-Markov model then reads

$$
\begin{aligned}
\tau_{obs}(t) = \quad & -\tfrac{1}{c}\,\mathbf{b}(t)\cdot\mathbf{W}(t)\cdot\mathbf{S}(t)\cdot\mathbf{Q}(t)\cdot\mathbf{k} \\
& -(\mathcal{T}_0^{(1)} + \mathcal{T}_1^{(1)}\cdot(t-t_0) + \mathcal{T}_2^{(1)}\cdot(t-t_0)^2) \\
& +(\mathcal{T}_0^{(2)} + \mathcal{T}_1^{(2)}\cdot(t-t_0) + \mathcal{T}_2^{(2)}\cdot(t-t_0)^2) \\
& -(T^{(1)}(t_0) + \tfrac{T^{(1)}(t_1)-T^{(1)}(t_0)}{t_1-t_0}(t-t_0) + \ldots + \tfrac{T^{(1)}(t_i)-T^{(1)}(t_{i-1})}{t_i-t_{i-1}}(t-t_i)) \\
& +(T^{(2)}(t_0) + \tfrac{T^{(2)}(t_1)-T^{(2)}(t_0)}{t_1-t_0}(t-t_0) + \ldots + \tfrac{T^{(2)}(t_i)-T^{(2)}(t_{i-1})}{t_i-t_{i-1}}(t-t_i)) \\
& +m_w^{(1)}\cdot[\mathcal{AT}^{(1)}(t_0) + \tfrac{\mathcal{AT}^{(1)}(t_1)-\mathcal{AT}^{(1)}(t_0)}{t_1-t_0}(t-t_0) + \ldots + \tfrac{\mathcal{AT}^{(1)}(t_i)-\mathcal{AT}^{(1)}(t_{i-1})}{t_i-t_{i-1}}(t-t_i)] \\
& +m_w^{(2)}\cdot[\mathcal{AT}^{(2)}(t_0) + \tfrac{\mathcal{AT}^{(2)}(t_1)-\mathcal{AT}^{(2)}(t_0)}{t_1-t_0}(t-t_0) + \ldots + \tfrac{\mathcal{AT}^{(2)}(t_i)-\mathcal{AT}^{(2)}(t_{i-1})}{t_i-t_{i-1}}(t-t_i)] \\
& +m_w^{(1)}(\varepsilon^{(1)})\cdot\cot\varepsilon^{(1)}\cdot[\mathcal{G}_n^{(1)}\cos\alpha^{(1)} + \mathcal{G}_e^{(1)}\sin\alpha^{(1)}] \\
& +m_w^{(2)}(\varepsilon^{(2)})\cdot\cot\varepsilon^{(2)}\cdot[\mathcal{G}_n^{(2)}\cos\alpha^{(2)} + \mathcal{G}_e^{(2)}\sin\alpha^{(2)}] \\
& +\tau_{corr} + \epsilon
\end{aligned}
\tag{13.114}
$$

with

$\mathscr{T}_j^*$	parameters of the clock polynomial at station 1 or 2
$T^{(*)}(t_i)$	coefficients of linear splines at telescope (1) or (2) for a specific epoch t_i parametrized as continuous piece-wise linear functions (CPWLF), i.e., linear splines
$m_w^{(*)}$	mapping function of the wet part of the atmosphere at telescope (1) or (2)
$\mathscr{AT}^*(t_i)$	zenith wet delays parametrized as CPWLF at telescope (1) or (2) for a specific epoch t_i
$\mathscr{G}_{n/e}^{(*)}$	atmospheric gradients in north or east direction at telescope (1) or (2)

It should be mentioned that in most analysis packages using least squares adjustments in Gauß-Markov models, the clock parameters are modelled with a second order polynomial and a superimposed piece-wise linear polygon as described above and formulated in Eq. 13.114. Although this seems to be an over-parameterization of the clock behavior and the correlation coefficient between the clock parameters are extremely high, the approach still serves a purpose. The reason is, that the course of the parameters in time can then be constrained more easily if the number of observations is insufficient (Sec. 13.7.5).

13.7.4. Partial derivatives

According to Eq. 13.96, the design matrix contains the partial derivatives of the observed delay with respect to the unknown parameters. Taking the extended observation equation, Eq. 13.114 as the basis, the partial derivatives can be sorted according to sequence of parameters as they appear in there.

The first parameters are the telescope coordinates embedded in the baseline vector according to Eq. 13.30. The **b** vector in turn appears in a product with a few other parameters, the EOP and radio source positions. The product rule of derivatives simplifies the following derivation because only the factor of the baseline vector is dependent on the telescope coordinates while the other factors are not. In addition, the same applies to all other lines of Eq. 13.114. For this reason, we can restrict the derivation of the partials w.r.t. the telescope coordinates to the geometry components in the first line of Eq. 13.114 which reads

$$\tau_{obs}(t) = -\frac{1}{c}\begin{pmatrix} x_{(2)} - x_{(1)} \\ y_{(2)} - y_{(1)} \\ z_{(2)} - z_{(1)} \end{pmatrix} \cdot \mathbf{W}(t) \cdot \mathbf{S}(t) \cdot \mathbf{Q}(t) \cdot \begin{pmatrix} \cos\delta_c \cdot \cos\alpha_c \\ \cos\delta_c \cdot \sin\alpha_c \\ \sin\delta_c \end{pmatrix} \tag{13.115}$$

when we expand the b and k matrices with the respective terrestrial coordinates according to Eq. 13.30 and celestial positions according to Eq. 13.31. The radio source positions α_c and δ_c are referred to the current realization of the ITRS, which are the most up-to-date catalog positions.

The partial derivatives, which we need, are those of the observed delay with respect to the parameter of interest at the time and location of the linarized geometry. In most cases the parameter is present only in one of the factors of Eq. 13.114. For this reason, the derivation of any partial derivative can be split into an affected and unaffected part such as

$$\frac{\partial \tau_{obs}(t)}{\partial x_{(1)}} = -\frac{1}{c}\frac{\partial \mathbf{b}}{\partial x_{(1)}} \cdot \mathbf{W}(t) \cdot \mathbf{S}(t) \cdot \mathbf{Q}(t) \cdot \mathbf{k} \tag{13.116}$$

reducing the computational costs considerably. In the case of the x coordinate of telescope (1), the derivative is

$$\frac{\partial \mathbf{b}}{\partial x_{(1)}} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}. \tag{13.117}$$

For the y and z components as well as for those of telescope (2), the matrix of derivatives looks correspondingly. The full expression for the x component of telescope (1), also as an example for all other parameters, then reads

$$\frac{\partial \tau_{obs}(t)}{\partial x_{(1)}} = -\frac{1}{c}\begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} \cdot \mathbf{W}(t) \cdot \mathbf{S}(t) \cdot \mathbf{Q}(t) \cdot \begin{pmatrix} \cos\delta_c \cdot \cos\alpha_c \\ \cos\delta_c \cdot \sin\alpha_c \\ \sin\delta_c \end{pmatrix} \tag{13.118}$$

For the derivations of the partial derivatives of the EOP, we need to recall the structure of the rotation matrices (Eqs. A.7 - A.9). The partial derivatives of these rotation matrices is achieved by differentiating all elements individually:

$$\frac{\partial \mathbf{R}_1(\alpha)}{\partial \alpha} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\sin\alpha & \cos\alpha \\ 0 & -\cos\alpha & -\sin\alpha \end{pmatrix} \tag{13.119}$$

$$\frac{\partial \mathbf{R}_2(\alpha)}{\partial \alpha} = \begin{pmatrix} -\sin\alpha & 0 & -\cos\alpha \\ 0 & 0 & 0 \\ \cos\alpha & 0 & -\sin\alpha \end{pmatrix} \tag{13.120}$$

$$\frac{\partial \mathbf{R}_3(\alpha)}{\partial \alpha} = \begin{pmatrix} -\sin\alpha & \cos\alpha & 0 \\ -\cos\alpha & -\sin\alpha & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{13.121}$$

For the partial derivatives for polar motion we can apply the same concept as for the telescope coordinates which in this case is just to perform the derivation of the respective rotation matrix in the model because none of the other factors depends on the polar motion parameters. This then yields

$$\frac{\partial \tau_{obs}(t)}{\partial (x_p|y_p)} = -\frac{1}{c}\mathbf{b} \cdot \frac{\partial \mathbf{W}(t)}{\partial (x_p|y_p)} \cdot \mathbf{S}(t) \cdot \mathbf{Q}(t) \cdot \mathbf{k} \tag{13.122}$$

Again, each element of the original rotation matrix, here it is Eq. 13.7, is to be deviated individually providing for the x_p component

$$\frac{\partial \mathbf{W}(t)}{\partial x_p} = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \tag{13.123}$$

and

$$\frac{\partial \mathbf{W}(t)}{\partial y_p} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \tag{13.124}$$

The multiplication in Eq. 13.122 then yields the numerical values. The same mechanism of

forming the partial derivatives just applying the derivation to the respective rotation matrix also applies to the *Earth rotation angle*:

$$\frac{\partial \tau_{obs}(t)}{\partial \Theta} = -\frac{1}{c}\mathbf{b} \cdot \mathbf{W}(t) \cdot \frac{\partial \mathbf{S}(t)}{\partial \Theta} \cdot \cdot \mathbf{Q}(t) \cdot \mathbf{k} \tag{13.125}$$

where the rotation matrix Eq. 13.12 is the basis for

$$\frac{\partial \mathbf{S}(t)}{\partial \Theta} = \begin{pmatrix} -\sin\Theta & \cos\Theta & 0 \\ -\cos\Theta & -\sin\Theta & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{13.126}$$

For the partial derivatives for the adjustments of the nutation offsets dX and dY we apply the same scheme as before with

$$\frac{\partial \tau_{obs}(t)}{\partial (X|Y)} = -\frac{1}{c}\mathbf{b} \cdot \mathbf{W}(t) \cdot \mathbf{S}(t) \cdot \frac{\partial \mathbf{Q}(t)}{\partial (X|Y)} \cdot \mathbf{k} \tag{13.127}$$

However, we have to recall that according to Eq. 13.26

$$\mathbf{Q}(t) = \mathbf{Q}'^{T}(t) \cdot \mathbf{R}_3(s) \tag{13.128}$$

both rotation matrices depend on X and Y. For this reason, we have to apply the product rule

$$\frac{\partial \mathbf{Q}(t)}{\partial X} = \frac{\partial \mathbf{Q}'^{T}(t)}{\partial X} \cdot \mathbf{R}_3(s) + \mathbf{Q}'^{T}(t) \cdot \frac{\partial \mathbf{R}_3(s)}{\partial X} \tag{13.129}$$

for the parameter X, and the same for Y

$$\frac{\partial \mathbf{Q}(t)}{\partial Y} = \frac{\partial \mathbf{Q}'^T(t)}{\partial Y} \cdot \mathbf{R}_3(s) + \mathbf{Q}'^T(t) \cdot \frac{\partial \mathbf{R}_3(s)}{\partial Y} \tag{13.130}$$

On the basis of the rotation matrix Eq. 13.22 together with Eq. 13.21 we find

$$\frac{\partial \mathbf{Q}'^T(t)}{\partial X} = \begin{pmatrix} -X - \frac{1}{2}X^3 - \frac{1}{4}XY^2 & -\frac{1}{2}Y - \frac{3}{8}X^2Y - \frac{1}{8}Y^3 & 1 \\ -\frac{1}{2}Y - \frac{3}{8}X^2Y - \frac{1}{8}Y^3 & -\frac{1}{4}XY^2 & 0 \\ -1 & 0 & -X - \frac{1}{2}X^3 - \frac{1}{2}XY^2 \end{pmatrix} \tag{13.131}$$

and

$$\frac{\partial \mathbf{Q}'^T(t)}{\partial Y} = \begin{pmatrix} -\frac{1}{4}X^2Y & -\frac{1}{2}X - \frac{3}{8}XY^2 - \frac{1}{8}X^3 & 0 \\ -\frac{1}{2}X - \frac{3}{8}XY^2 - \frac{1}{8}X^3 & -Y - \frac{1}{4}X^2Y - \frac{1}{2}Y^3 & 1 \\ 0 & -1 & -Y - \frac{1}{2}Y^3 - \frac{1}{2}X^2Y \end{pmatrix} \tag{13.132}$$

and for the CIO locator s according to Eq. 13.25 and the respective rotation matrix around the 3-axis:

$$\frac{\partial \mathbf{R}_3(s)}{\partial X} = \begin{pmatrix} \frac{Y}{2}\sin s & -\frac{Y}{2}\cos s & 0 \\ \frac{Y}{2}\cos s & \frac{Y}{2}\sin s & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{13.133}$$

and

$$\frac{\partial \mathbf{R}_3(s)}{\partial Y} = \begin{pmatrix} \frac{X}{2}\sin s & -\frac{X}{2}\cos s & 0 \\ \frac{X}{2}\cos s & \frac{X}{2}\sin s & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{13.134}$$

The final geometrical parameters are the position components of the source positions where we apply

$$\frac{\partial \tau_{obs}(t)}{\partial(\alpha|\delta)} = -\frac{1}{c}\,\mathbf{b} \cdot \mathbf{W}(t) \cdot \mathbf{S}(t) \cdot \mathbf{Q}(t) \cdot \frac{\partial \mathbf{k}}{\partial(\alpha|\delta)} \tag{13.135}$$

Performing the partial differentiation on Eq. 13.31 yields

$$\frac{\partial \tau_{obs}(t)}{\partial \alpha} = -\frac{1}{c}\,\mathbf{b} \cdot \mathbf{W}(t) \cdot \mathbf{S}(t) \cdot \mathbf{Q}(t) \cdot \begin{pmatrix} -\cos\delta_c \cdot \sin\alpha_c \\ \cos\delta_c \cdot \cos\alpha_c \\ 0 \end{pmatrix} \tag{13.136}$$

$$\frac{\partial \tau_{obs}(t)}{\partial \alpha} = -\frac{1}{c}\,\mathbf{b} \cdot \mathbf{W}(t) \cdot \mathbf{S}(t) \cdot \mathbf{Q}(t) \cdot \begin{pmatrix} -\sin\delta_c \cdot \cos\alpha_c \\ -\sin\delta_c \cdot \sin\alpha_c \\ \cos\delta_c \end{pmatrix} \tag{13.137}$$

Besides the purely geometric parameters we also have to estimate clock and atmosphere parameters. The partial derivatives for both groups again follow the observation equation Eq. 13.114. For the clock offsets, rates and second order terms we find

$$\frac{\partial \tau_{obs}(t)}{\partial \mathcal{T}_0^{(*)}} = 1 \qquad \frac{\partial \tau_{obs}(t)}{\partial \mathcal{T}_1^{(*)}} = (t - t_0) \qquad \frac{\partial \tau_{obs}(t)}{\partial \mathcal{T}_2^{(*)}} = (t - t_0)^2 \tag{13.138}$$

$(*)$ is again the wildcard index for the telescopes.

The partials for the piece-wise linear offsets as formulated in Eq. 13.112 are a bit more complicated to organize because they depend on the interval boundaries. They have the form

$$\begin{aligned}
\frac{\partial \tau_{obs}(t)}{\partial \mathcal{T}_{i-1}} &= \begin{cases} 1 - \frac{t-t_{i-1}}{t_i-t_{i-1}} & \text{for } t_{i-1} < t < t_i \\ 0 & \text{for all other epochs} \end{cases} \\
\frac{\partial \tau_{obs}(t)}{\partial \mathcal{T}_{i}} &= \begin{cases} \frac{t-t_{i-1}}{t_i-t_{i-1}} & \text{for } t_{i-1} < t < t_i \\ 0 & \text{for all other epochs} \end{cases}
\end{aligned} \tag{13.139}$$

This leads to a block diagonal scheme of the partials in the design matrix as depicted in Fig. 13.12.

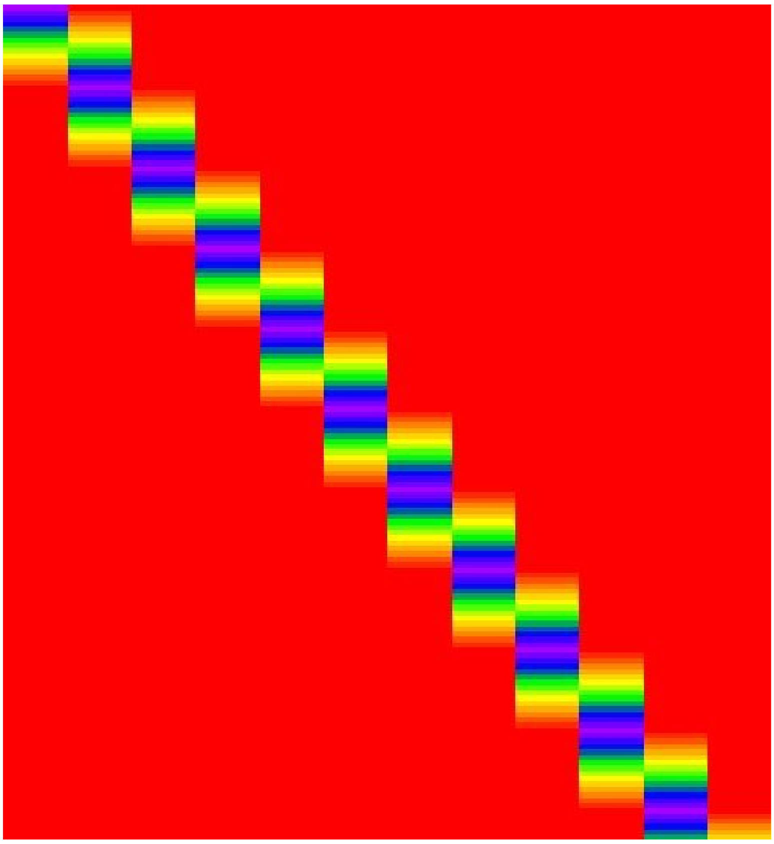


Figure 13.12: Schematic distribution of magnitudes of partial derivatives for piece-wise linear offsets, lines for observations, columns for parameters (red = 0, blue = 1)

Finally, we come to the last group of standard parameters, the (excess) zenith wet path delays. As for the clock parameters, the simplest approach is to estimate just one offset for the whole session but this, of course, is not sufficient, not even with a temporal slope, because the atmosphere is highly variable in time and space. As a first refined approach, the session interval is subdivided into multiple piece-wise linear segments of regular duration between 15 minutes and a few hours. Here, the same applies as with the clock parameters namely that according to Eq. 13.139 the observations contribute information to the parameters only within their segment. For the general ZWD offset, the partial derivative for telescope $(*)$ is simply the mapping function (as introduced in Eq. 13.56)

$$\frac{\partial \tau_{obs}(t)}{\partial \mathcal{AT}^{(*)}(t_0)} = m_w^{(*)}(\varepsilon) \tag{13.140}$$

As for the clock parameters, we can again increase the temporal resolution by introducing piece-wise linear segments for which the offsets at the beginning and end of each segment are set up for estimation. The partial derivatives are again derived from Eq. 13.114 following

$$\begin{aligned}
\frac{\partial \tau_{obs}(t)}{\partial \mathcal{AT}_{i-1}} &= \begin{cases} m_w^{(1)}(\varepsilon) \cdot (1 - \frac{t-t_{i-1}}{t_i-t_{i-1}}) & \text{for } t_{i-1} < t < t_i \\ 0 & \text{for all other epochs} \end{cases} \\
\frac{\partial \tau_{obs}(t)}{\partial \mathcal{AT}_{i}} &= \begin{cases} m_w^{(1)}(\varepsilon) \cdot (\frac{t-t_{i-1}}{t_i-t_{i-1}}) & \text{for } t_{i} < t < t_{i+1} \\ 0 & \text{for all other epochs} \end{cases}
\end{aligned} \tag{13.141}$$

The use of continuous piece-wise linear offset parameter estimation is not restricted to clocks and zenith wet delays alone but can also be applied to other target parameters such as the Earth rotation parameters. The general formulation for the partial derivatives for a parameter $\tilde{x}$ is

$$\begin{aligned}
\frac{\partial \tau_{obs}(t)}{\partial \tilde{x}_{i-1}} &= \begin{cases} \frac{\partial \tau_{obs}(t)}{\partial \tilde{x}} \cdot (1 - \frac{t-t_{i-1}}{t_i-t_{i-1}}) & \text{for } t_{i-1} < t < t_i \\ 0 & \text{for all other epochs} \end{cases} \\
\frac{\partial \tau_{obs}(t)}{\partial \tilde{x}_{i}} &= \begin{cases} \frac{\partial \tau_{obs}(t)}{\partial \tilde{x}} \cdot (\frac{t-t_{i-1}}{t_i-t_{i-1}}) & \text{for } t_{i-1} < t < t_i \\ 0 & \text{for all other epochs} \end{cases}
\end{aligned} \tag{13.142}$$

13.7.5. Constraining non-datum parameters

Due to the fact that sometimes an insufficient number of observations is available or a general deficit in the geometric stability of the normal equation system exists, the normal equation system may have to be augmented with constraints for certain parameters. This applies predominantly to parameters which are introduced to enhance the time resolution through continuous piece-wise linear functions as applied to clock and atmosphere but also to Earth orientation parameters (Sec. 13.7.3). The constraints are mostly formulated as pseudo-observations and serve to stabilize timely adjacent parameters reducing their erratic variability originating from lack of observations. This section is called "Constraining non-datum parameters" because for datum definition a few more issues have to be taken into account (Secs. 13.8 and 13.9.1).

Constraints for non-datum parameters mainly determine that two adjacent parameters are identical within a certain standard deviation. On the basis of Eq. 13.112, we can write an observation equation in the form of

$$f(t_j) - f(t_{j-1}) = 0. \tag{13.143}$$

Assigned to this is a freely scalable standard deviation which determines how strong the constraint acts. Small errors induce high weights of the constraint equations and thus strong constraints, large errors produce weak constraints. The standard deviation depends on the parameters to be constrained. Of course, it should not be chosen too small in order not to dominate the estimate of the respective parameters because then the difference in Eq. 13.143 is forced to zero. For zenith wet path delays common practice is to assign around 50 ps (15 mm) and for clock parameters around 40 ps (12 mm) for the respective standard deviations σ_c.

The linearization of the observation equation Eq. 13.143 produces two partial derivatives with the values of 1 for the parameter at $f(t_j)$ and -1 for the preceding parameter at $f(t_{j-1})$. For m parameters $f(t_j)$, we can write $m-1$ differences. Ordering these in a design matrix of the constraining pseudo-observations $\mathbf{C}_{m-1,m}$ yields

$$\mathbf{C}_{m-1,m-1} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}. \tag{13.144}$$

The inverse of the covariance matrix $\mathbf{\Sigma}_{cc}^{-1}$, which is the weight matrix of the constraints $\mathbf{P}_{cc}$, is

$$\mathbf{P}_{cc} = \mathbf{\Sigma}_{\mathbf{cc}}^{-1} = \begin{pmatrix} 1/\sigma_c^2 & 0 & \dots & 0 \\ 0 & 1/\sigma_c^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\sigma_c^2 \end{pmatrix} = \begin{pmatrix} p_{11} & 0 & \dots & 0 \\ 0 & p_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p_{m-1,m-1} \end{pmatrix}. \tag{13.145}$$

To apply the constraints to the basic normal equation system $\mathbf{N}_o$, we also have to compute a normal matrix of the constraint equations. This is

$$\mathbf{N}_c = \mathbf{C}^T \mathbf{\Sigma}_{cc}^{-1} \mathbf{C} \tag{13.146}$$

For m parameters, the dimension of $\mathbf{N}_c$ is (m x m) = (m x m-1) · (m-1 x m-1) · (m-1 x m). The right hand side of the normal equation system of the constraints is zero. The composition of the constrained normal matrix can be performed by adding the original normal matrix N_o and the datum matrix N_c

$$\mathbf{N}_c = \mathbf{N}_o + \mathbf{N}_c = \mathbf{A}^T \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{A} + \mathbf{C}^T \boldsymbol{\Sigma}_{cc}^{-1} \mathbf{C} \tag{13.147}$$

Special care has to be taken that the positions of the elements of the normal matrix of constraints match those of the original normal matrix.

13.7.6. Phase observations

Besides the group delay observations, which are the principle working domain of the VLBI analysts, the interferometer phase observations are of interest in selected cases as well. One area of applications for the phase observables are local, short distance VLBI observing sessions. As already described in Sec. 2.2, the phase observable may come along as a total phase or as a phase delay. Mostly, the fringe fitting programs such as *fourfit* or *PIMA* report them in station-based time reference frame. In principle, for these the same applies as for the group delay observations as described above, i.e., the geophysical models and the estimation process are the same, for example, as in Eq. 13.114. However, we need to take into account a few peculiarities of the phase observables.

The first effect to be taken care of has to do with the fact that the feed horns of telescopes with azimuth-elevation (Fig. 5.3) or XY mounts (Fig. 5.5) change their principal axes with respect to the celestial sphere when switching between radio sources. The effect applies to both circularly and linearly polarized feed horns. By the way, this is not the case for polar mounts (Fig. 5.4).

For both telescopes (i), the parallactic angles $\psi_{RS}(i)$ need to be computed (see Appendix B) and to be subtracted for the baseline

$$\Delta\psi_{RS} = \psi_{RS}^{(2)} - \psi_{RS}^{(1)}. \tag{13.148}$$

For application to the phase delay (as a time delay quantity) it has to be divided by $2\pi \cdot \nu_0$ (Cotton, 1993)

$$\Delta\tau_{\psi_{RS}} = \frac{\psi_{RS}^{(2)} - \psi_{RS}^{(1)}}{2\pi \cdot \nu_0} \tag{13.149}$$

ψ_{RS} is zero for observations along the local meridian serving as the reference for applying corrections to the observed phase for feed rotations.

The second effect is the ionosphere of generally all dispersive effects. Although the ionosphere induces the need for corrections only for baselines longer than a few tens of kilometers, the ionosphere calibrations needs to be mentioned here for completeness. The reason is that comparing Eqs. 13.70 and 13.67, we see that the sign of the correction for group delays and phases is opposite. For this reason, we need to be careful with the sign when applying total phase and phase delay corrections from dual band observations. Unfortunately, even the ionosphere calibrations suffer from the 2π ambiguity problem and attempts to use phase observations for longer baselines were only tried many years ago (Petrov, 1999).

Depending on the analysis software, phase observations can either be handled in the same way as group delays, i.e., using the phase delay (in seconds), or as phase observable using total phase (in radian) (Sec. 11.6) provided that the above is taken into account. The method of choice is mostly the use of the phase delay. According to Eq. 11.9 the initial representation of this quantity contains the apriori (model) delay of the correlator and is thus free of the many phase turn ambiguities already "known" to the model.

The actual way of handling the observed total phases or phase delays are handled in the Level 2 analysis stage depends on how the observables are carried over in the analysis pipeline. At the time of writing (August 2021), only the total phase computed according to Eq. 11.11 is carried over from *fourfit* via *vgosDBmake* to the vgosDB data bases. This requires additional computing steps to derive the phase delay for the analysis programs.

First, the phase delay ambiguities need to be determined. The ambiguity spacing $\square\tau_{ph}^{amb}$ is inferred from the observing frequency (reference frequency) ν_0 in Hz with

$$\square\tau_{ph}^{amb} = \frac{1}{\nu_0} \tag{13.150}$$

which is around 100 - 120 ps at X band.

The residual phase delay $\Delta\tau_{ph}$ in seconds determined from the ambiguous total phase (mod 2π) $\phi_{total}(\nu_0)$ in radians as exported from the fringe fitting program is

$$\Delta\tau_{ph} = \frac{\phi_{total}(\nu_0)}{2\pi \cdot \nu_0} \tag{13.151}$$

N.B.: *fourfit* exports $\phi_{total}(\nu_0)$ in degrees.

Then, the number of phase ambiguities N can be computed in multiple ways depending on what composes the delay which is represented by N, τ_N. If we want to reproduce what the fringe fitting program, e.g., *fourfit*, subtracted initially, we have to apply the a priori delay of the correlator model τ_g and subtract the ambiguous total phase delay $\Delta\tau_{ph}$ (compare Eq. 11.11 and 11.12) and apply the delay contribution of the parallactic angle difference $\Delta\tau_{\psi_{RS}}$

$$\tau_{N_1} = \tau_g - \Delta\tau_{ph} - \Delta\tau_{\psi_{RS}}. \tag{13.152}$$

We also have the option to use the observed total group delay τ_{gr} and to add any multiple n_{gr} of the group delay ambiguity spacing $\square\tau_{gr}^{amb}$ (mostly 50, 100 or 200 ns) originating from the FFT process in fringe fitting and identified in the analysis process

$$\tau_{N_2} = \tau_{gr} - \Delta\tau_{ph} - \Delta\tau_{\psi_{RS}} [+n_{gr} \cdot \square\tau_{gr}^{amb}] \tag{13.153}$$

The residual phase delay $\Delta\tau_{ph}$ is subtracted again for compensating for Eq. 11.11 as in Eq. 13.152.

The integer number of phase delay ambiguities N can then be computed with

$$N = round\left[\frac{\tau_N}{\square\tau_{ph}^{amb}}\right] \tag{13.154}$$

where τ_N is one of the two options of N as above and $\square\tau_{ph}^{amb}$ is the phase delay ambiguity spacing of Eq. 13.150. The rounding is done to make N an integer value. If N_1 is being used, the rounding should not even be needed because Eq. 13.154 just reproduces what *fourfit* does with the modulo 2π functions.

Finally, the (total) phase delay to be used in the analysis program is computed as

$$\tau_{ph} = \Delta\tau_{ph} + N \cdot \Box\tau_{ph}^{amb} \tag{13.155}$$

[These equations following N_2 represent what is done within *nuSolve* and it should be pointed out that the parallactic angle correction has been applied (already).]

Obviously, Eq. 13.154 forces the ambiguities to be integer and applies the resulting integer multiples to the total phase delay in Eq. 13.155. As described above, τ_N can be computed in different ways and consequently provide different results. However, as empirical tests have shown, the difference between the two total phase delay computations (either with N_1 or N_2) is always an integer number of ambiguity spacings $\Box\tau_{ph}^{amb}$. The reason is that the sum of the roundings for the two different N exactly matches the rounding of the $\bar{N}$ for the observed minus a priori delay difference, sometimes including a respective ± 1 integer. So, the difference between the two paths for N and the respective phase delay is uncritical because the final phase delay ambiguities need to be identified and compensated for in a separate analysis step (Herring, 1983).

Finally, we have to look whether the partial derivatives of the observations with respect to the parameters need to be modified. Since τ_{ph} is a time delay, we can use the partials of Sec. 13.7.4 directly. For the total phases, we use sequential derivation for the parameters P and find

$$\frac{\partial \phi_{total}}{\partial P} = \frac{\partial \phi_{total}}{\partial \tau_{ph}} \cdot \frac{\partial \tau_{ph}}{\partial P} \tag{13.156}$$

Taking Eq. 13.151 and reorder it to solve for ϕ_{total} we can do the first part of the differentiation and find

$$\frac{\partial \phi_{total}}{\partial \tau_{ph}} = 2\pi \cdot \nu_0 \tag{13.157}$$

which is a constant. So, in case we would work with the phase angles, we just have to multiply the partials from Sec. 13.7.4 with $2\pi \cdot \nu_0$.

As stated already, phase delay solutions were successfully applied to observations of very short local baselines (Hase and Petrov, 1999; Niell et al., 2021; Varenius et al., 2021). Here, only X band observations were used. For phase delay observations of longer baselines, where ionosphere calibration is necessary, Petrov (1999) devised an analysis scheme. Unfortunately, system stability issues inhibited further progress at that time.

13.7.7. Solution types

Before turning to the final estimation, we have to distinguish between single session analysis and multi-session, so-called global, analysis. The standard parameter set for single session solutions consists of four groups of parameters which are considered as local parameters, i.e., only valid at the epoch of the session:

- the parameters of the clock model
- radio telescope coordinates x_i, y_i, z_i
- Earth orientation parameters, i.e., the two polar motion components x_p, y_p, the Earth's phase of rotation UT1–UTC, two adjustment parameters for the precession/nutation model dX, dY
- the parameters of the refraction model due to the wet component of the atmosphere.

The single session analysis always is the first step which is mainly devoted to preparing the session data for establishing the final setup of this session. First of all, the ambiguities have to be resolved, which result from possible ambiguous peak identification due to inaccurate a priori information of geometry and clocks (Sec. 11.5). Special care has to be taken that in all possible triangles the sum of the delays has to close to zero within a few ns.

After this has been done, the X band group delays are corrected for ionospheric refraction effects according to Eq. 13.71. It can be expected that systematic biases between the S and X band delays prevail which contaminate the ionosphere correction. However, since these are of a constant nature, they will finally disappear in the clock parameter estimates.

Next comes the parameterization of the clock and atmosphere parameters mainly depending on the number of individual delay observations and the possible and necessary time resolution of the parameters. Eq. 13.114 contains all the parameters which are estimated in routine processing. However, we have to consider two peculiarities. The first is that we can only estimate clock parameters which are relative to a reference clock. Regularly one of the most stable clocks is selected from experience. For this purpose the design matrix does not contain coefficients for the reference clock resolving the singularity which would appear otherwise.

The second caveat is the estimation of the atmospheric refraction parameters. Normally, VLBI telescopes are far enough apart that the estimates of the wet zenith path delays of any two telescopes are de-correlated. However, with the advent of more than one radio telescope at a single observatory, e.g., at Wettzell (Germany), Onsala (Sweden), or Ny Alesund (Norway), the partial derivatives of the first order wet zenith delay of two nearby telescopes according to Eq. 13.140 and of course all other continuous piece-wise linear polygons are highly correlated through the almost identical elevation angles. For this reason, only relative zenith wet delays can be estimated in cases where the two telescopes are close together. As soon as a third telescope provides observations on longer baselines with different elevation angles, the estimation of absolute zenith wet delays for all telescopes becomes possible again.

The estimation of parameters of piecewise linear functions does only work if a sufficient number of observations is available in each segment. Unfortunately, sometimes this is not the case because of failures or deliberate exclusions. In such a case the system of normal equations would become degenerate. However, this is regularly cured by introducing constraints in the form of pseudo observations to the effect that the rates between every two functional values should be zero

$$\frac{d(F(t_i) - F(t_{i-1}))}{dt} = 0. \tag{13.158}$$

The selection of the variances of these observations determines how loosely or how tightly the constraint affects the estimation. For the clock parameters normally a standard deviation on the order of about 5×10^{-14} s/s and for the zenith wet delay of about 50 ps/h is being chosen.

Another issue concerns the fact that the VLBI technique is only capable of producing relative telescope coordinates. This means that any solution setup per se is free of datum and the normal equation system would be singular if all telescope coordinates and the Earth rotation parameters (x_p, y_p, UT1–UTC) are tried to be estimated. It has a rank defect of six, three translations and three rotations. In a single session, the simplest option would be to fix the three coordinate components of one telescope and the three Earth rotation parameters, i.e., eliminate them from the normal equation system (Sec. 13.8.2).

Another more flexible way being less dependent on the coordinates of a single telescope is the introduction of no-net-rotation (NNR) and no-net-translation (NNT) conditions for at least three non-collinear telescopes or even all of them in the network (Angermann et al., 2004) (Sec. 13.8.3). The selection of the datum telescopes depends on quality of the a priori coordinates being suitable for the datum definition at hand.

As a final step of the session-wise analysis, an iterative identification process of significant outliers in the observations can be added applying different kinds of outlier identification schemes (Koch, 1999). Thus outliers can then be down-weighted for exclusion from the parameter estimation.

The second type of solutions is a multi-session or global solution. These solutions are carried out to determine parameters which are valid/constant for the whole observing period. They serve to estimate

- radio source positions
- telescope coordinates at a reference epoch
- linear or higher order velocities of the radio telescopes due to geodynamic effects

Other experimental solutions can be conceived, e.g., for the estimation of the relativistic factor γ or Love numbers.

A simple way of running these solutions is by stacking pre-reduced normal equation systems from the session-wise analysis. Pre-reduction in this context means that parameters are excluded from the solution vector by reducing the system of normal equations by

Gaußian elimination steps (Angermann et al., 2004). When all local parameters are pre-reduced from the session's normal equation system, it contains only the coefficients of the global parameters set up initially. New lines and columns, have to be entered for the linear or otherwise parameterized, e.g., harmonic, telescope motions (Angermann et al., 2004).

The radio source positions are mostly treated either as constant global parameters or as session-wise, so-called arc, parameters. The latter approach is chosen when a radio source evidently manifests position variations. However, new approaches also allow for intermediate modes with linear spline approximations at predetermined or automatically chosen time intervals (Karbon et al., 2017).

As the final step, the global parameters, as determined by the stacked normal equation system, are reinserted in the original session normal equation systems to also estimate the local parameters. With all parameters estimated the observation residuals can be computed and an assessment of the errors can be carried out.

The general concept of treating the datum in global solutions is the similar to that of single session solutions with NNR/NNT conditions (Angermann et al., 2004)(Sec. 13.8.1). In most of the cases, the latest realization of the ITRF (Altamimi et al., 2016) is taken as the datum frame. For the radio source positions, mostly the defining sources of the ICRF3 (Charlot et al., 2020) are chosen as the celestial datum for new and/or improved radio source positions.

13.8. Terrestrial datum definitions

13.8.1. General and functional considerations of terrestrial datum definition

Geodetic and astrometric VLBI data analysis, as for any other geodetic technique, is at first based on relative relationships in-between the observing network. A group of geodetic points, i.e., telescope reference points on Earth, in three-dimensional space at a specific epoch, is a pure point cloud. In the context of geodesy, the ensemble of telescope coordinates can be considered as a (survey) net or likewise as a spatial polyhedron with points at or near the Earth's surface (Fig. 13.13).

The VLBI observations, which we perform, establish very precise relative relationships between these points. Sometimes these relationships are called the geometric *configuration*. While the configuration is determined by the observations, there is no such thing for the absolute position of the polyhedron or network. For this, we apply a datum with a datum definition. Under the expression *datum* we understand all definitions which are needed to locate a stiff polyhedron at a certain place in space with a specific orientation of its axes. In general, datum definition is a purely conventional (scientifically agreed upon) fixing where it was/is tried to come close to a physical or at least geometric framework as much as possible.

On Earth, the VLBI network has six degrees of freedom, three translations and three rotations. In a 3D Cartesian system, the three translations define where the origin of the

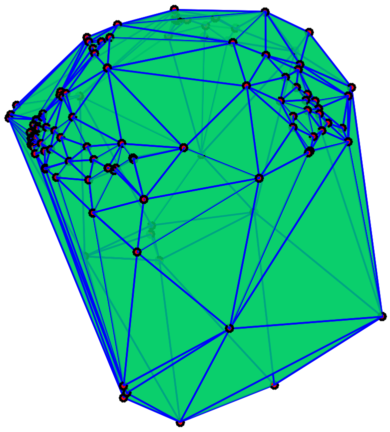


Figure 13.13: Datum-free polyhedron of past and current VLBI telescopes. Europe is in the top right, North America top left (Courtesy: Armin Corbin).

coordinates is, the three rotations define in which directions the axes point. In other words, we may shift our VLBI polyhedron along the coordinate axes (translations) and rotate the network around the same axes (rotations) into a target frame. The scale, which is the seventh parameter of a 3D similarity transformation, is fixed explicitly in the data analysis through fixing the speed of light to the conventional value (299,792,458 m/s).

In the context of datum definition, we also have to take into account that the locations of the radio telescopes are time dependent because tectonics causes motions of the lithosphere. To first order, these motions are linear, which means that, in the simplest approach but sufficient for the purpose at hand, we can describe them as time derivatives or velocities. The coordinates then need a time tag and the coordinates $\mathbf{x_t}$ of a telescope at a certain epoch t need to be computed as a composite of the coordinates $\mathbf{x_{t_0}}$ at a reference epoch t_0 and the velocity $\mathbf{v}$ "integrated" over the time elapsed since the reference epoch $t - t_0$ (Eq. 13.159).

$$\mathbf{x}_t = \mathbf{x}_{t_0} + \mathbf{v}_x \cdot (t - t_0) \tag{13.159}$$

As a consequence of this, any advanced datum definition also needs a special treatment of the velocities (see below).

Datum definition can be done in multiple ways as part of the data analysis of VLBI observations, i.e., within the adjustment process. In almost all cases, the datum is applied by using a conventional coordinate frame as the target frame.

As already explained above, VLBI observations produce (only) relative configurations of telescopes. The task of datum definition now is to define the location and orientation of the free network or polyhedron in a conventional frame to be able to produce coordinates in a pre-defined frame. Examples of such conventional frames were described in Sec. 13.1. The need for this may also arise from the facts that the configuration of the polyhedron of VLBI telescopes may be slightly different to that of the conventional frame (because of new and better observations and analysis models) and/or because there may be new stations which shall be mapped into a conventional frame. Here, we address the mathematical background of datum definition in the context of VLBI with its six degrees of freedom for the terrestrial frame.

When we consider the functional models for datum definition, we have two options. In the first case, telescope coordinates and their covariances have already been determined in an arbitrary reference frame in a previous VLBI analysis process. We then apply a similarity transformation to shift and rotate this frame into a conventional one. This can be done in different ways. In the second case, we incorporate the functionality of the first step in our VLBI analysis software and apply the datum directly in the analysis. This has the advantage that we save a processing step but may blank out the transformation parameters which sometimes are of interest as well.

Although the word *transformation* suggests that we change the form of the polyhedron or network, we only consider rigid (similarity) transformations where the polyhedron or network, respectively its configuration, is not distorted. The form is determined purely by the VLBI observations.

Before we elaborate on the different methods of datum definition, we should sketch the general steps and approaches. Let us assume that we have formulated the geometric model of the datum in some way or other. Then we need to apply it to our VLBI adjustment problem. In general, we can either apply the datum to the design/Jacobi matrix of the VLBI adjustment or we apply it to the normal matrix (Tab. 13.3). This depends on the realization in the individual software packages. In the case of applying the datum to the design matrix, pseudo-observations are added which reflect information on the datum. For multi-session analyses, of course the fact, that the telescope coordinates are time dependent, needs to be formulated for each session individually.

Table 13.3: Methods of terrestrial datum definition.

	design/Jacobi matrix	Normal matrix		
single session datum	pseudo observations	3-2-1 rule	Constraint matrix $\mathbf{B}^\mathbf{T}\mathbf{\Sigma}_{\mathbf{dd}}^{-1}\mathbf{B}$	Helmert rendering
multi session datum	pseudo observations incl. vel.	3-2-1 rule incl. vel.	Constraint matrix $\mathbf{B}^\mathbf{T}\mathbf{\Sigma}_{\mathbf{dd}}^{-1}\mathbf{B}$ incl. vel.	Helmert rendering incl. vel.

Application of the datum to the normal equation system is the most common and practical way because it can be applied to single- and multi-session solutions alike. Especially for multi-session solutions, the datum can be applied at the end, i.e., for the whole suit of sessions as the last step after stacking all normal matrices.

Before we address the different methods of how to apply the datum to the normal equation system, we should state that two expressions are in use, conditions and constraints.

Conditions force the geometric model onto the VLBI configuration while the strength of constraints can be controlled by formal errors. Conditions can be imposed by the 3-2-1 terrestrial datum rule (Sec. 13.8.2) or by rendering the normal equation system with Helmert parameters (Sec. 13.8.6). In both cases, there is no need for a covariance matrix of the conditions because they have quasi infinite weight. In the case of constraints, we need to set up and populate a covariance matrix of constraints. Of course, conditions can always be applied to act like constraints if very small formal errors are introduced (e.g., 10^{-3} mm) resulting in extreme weights. In this respect, conditions are only a special case of constraints. For this reason, sometimes in publications and especially in presentations, there is no clear distinction being made. The most common way of datum definition today is the no-net-translation and no-net-rotation constraint (Sec. 13.8.3) which considers a defined ensemble of the points in 3D space to match the conventional target frame as best as possible.

13.8.2. 3-2-1 terrestrial datum rule

For a better understanding of the more complicated procedures of datum definition, we start with a simple approach which has in fact been in use at the early days of geodetic VLBI. To recall, a VLBI solution has six degrees of freedom or a rank defect of six. We use the three translational degrees to define how far away our telescope coordinates are from the origin. The three rotational degrees are used to define the directions of the coordinate axes. This information is all embedded in the telescope coordinates. For this reason, we can just apply our definitions to some of the coordinates alone. The clue is the 3-2-1 rule. We take one telescope and assign the full set of three coordinate components, e.g., from the ITRF2020 table, to this. Then we take a second telescope somewhere else on Earth and assign just two coordinate components to this, and finally just one component of the third telescope is fixed. It is quite obvious that the three telescopes used for this datum definition should be rather far away from each other and should not form a great circle.

The 3-2-1 rule can be applied "externally" if two sets of coordinates are available already (Appendix D). However, and this is the more common approach, it can also be adopted in the least-squares analysis "internally". We can perform the fixing of the six coordinate components by eliminating the respective columns in the Jacobi matrix or deleting the respective rows and columns of the normal equation system leading to an implicit datum definition. As already stated, over longer periods of time we also have to take into account the velocities of the telescopes. Then the six velocity components of the same telescopes have to be treated accordingly, i.e., the respective rows and columns of the normal equation system need to be eliminated. The method is called a 'minimally constrained datum definition'.

In the early days of NASA's Crustal Dynamics Project (CDP), when the number of telescopes was still rather small, this datum definition was applied with the following extension.

Since the data sets started to cover more and more years, tectonic motion/continental drift had to be taken into account as well (Ma et al., 1989). This complication was solved by using the model velocity AM0-2 (Minster and Jordan, 1978) for the three components of the telescope WESTFORD (Massachusetts, USA) and the two purely horizontal components of RICHMOND (Florida, USA). The 3-2-1 datum rule was completed by setting the vertical velocity of KOKEE (Hawaii, USA) to zero. With this datum definition for the telescope's velocities, the other velocity components could be estimated as well (by including the datum definition in the analysis software directly). The deficit of this approach, however, was and still is that the uncertainties of one or more of the components, which are fixed, directly affect all other components to be estimated. For this reason, this procedure is not used in operational analyses any more but was described here for didactic purposes nevertheless because sometimes it comes quite handy for special investigations.

In the same category fall the following considerations as well. The datum definition as described here and also later always allow to also estimate the three Earth rotation parameters (ERP), polar motion x_p and y_p, and the Earth's phase of rotation UT1-UTC. The 3-2-1 rule can also be used in the way that the three components of one telescope are fixed (3 translations) and the three ERP of a session are fixed to some tabulated values (e.g., IERS C04 series)[25] providing the three rotation definitions for the full datum. Doing this, all other telescope coordinates can be estimated.

13.8.3. Terrestrial datum definition with no-net-translation and no-net-rotation constraints

Today, the no-net-translation (NNT) and no-net-rotation (NNR) approach is the preferred method for datum definition. With this method, a whole set of several (new) radio telescope coordinates, and often of their velocities, is mapped onto a conventional (old) set. The choice of several datum transfer telescopes avoids that the erroneous coordinates of a single telescope have a direct 1:1 impact. The more sets of reliable telescope coordinates are used, the smaller the probability that an outlier remains undetected. When selecting the telescopes for datum transfer, one should always look at the residuals for identification of possible outliers.

At the beginning of this section, we look at the NNR/NNT condition as such and its formalistic derivation. The result will be a condition or constraint matrix **B**. The use of this matrix provides several options which will be looked at at the end.

The NNR/NNT method originates from describing velocity fields within the Earth and on the Earth's surface. To avoid interactions of the movements of continental plates or arrays of geodetic observatories and 3D Earth rotation variability in their mathematical description, a Tisserand system is defined for the surface of the Earth. The axes of the

[25] https://www.iers.org/IERS/EN/DataProducts/EarthOrientationData/eop.html

Tisserand reference system have to obey the conditions

$$\int_E \mathbf{v}' \, dm = \mathbf{0} \qquad \text{No translation condition} \tag{13.160}$$

$$\int_E \mathbf{r} \times \mathbf{v}' \, dm = \vec{0} \qquad \text{No rotation condition} \tag{13.161}$$

with $\mathbf{v}'$ being the velocity field derived from a geophysical model and $\mathbf{r}$ being the localization (position) vector. Initially, the conditions should cover the entire Earth volume, thus the integration over the whole Earth E with mass elements m. With the hypothesis of a spherical Earth with uniform density, the volume integral becomes a surface integral. Taken over to the kinematic surface velocity problem, the rotation condition can also be interpreted as the total angular momentum of the whole surface S integrated over the area A being constrained to sum up to zero:

$$\int_S \mathbf{r} \times \mathbf{v}' \, dA = \mathbf{0} \tag{13.162}$$

In global frame investigations based on geophysical data such as the NNR-NUVEL-1A model (Argus and Gordon, 1991; Demets et al., 1990), the NNR condition separates individual plates first and sums them up

$$\int_S \mathbf{r} \times \mathbf{v}' \, dA = \sum_P \int_P \mathbf{r} \times \mathbf{v}' \, dA. \tag{13.163}$$

Setting up of a B^T matrix

In the context of datum definition, the velocities of the geophysical models $\mathbf{v}'$ can be replaced by the vectorial residuals of the transformation from one frame $\mathbf{x}$ to the other $\tilde{\mathbf{x}}$ with $\Delta\mathbf{x} = \mathbf{x} - \tilde{\mathbf{x}}$. Since we deal with individual radio telescopes, Eqs. 13.160 and 13.161 convert to

$$\sum_{i=1}^{N} \mathbf{\Delta x}^{(i)} = \mathbf{0} \qquad \text{No-net-translation constraint} \tag{13.164}$$

and

$$\sum_{i=1}^{N} (\mathbf{r}^{(i)} \times \mathbf{\Delta x}^{(i)}) = \mathbf{0} \qquad \text{No-net-rotation constraint,} \tag{13.165}$$

respectively, with N telescopes selected for the datum definition. The word net was introduced to indicate that the constraints are meant as net values being sums of arguments. In vector component writing, the NNT constraints of the coordinates are

$$\sum_{i=1}^{N} \mathbf{\Delta x}^{(i)} = \sum_{i=1}^{N} \begin{pmatrix} \Delta x^{(i)} \\ \Delta y^{(i)} \\ \Delta z^{(i)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{13.166}$$

The arguments of the sums can be re-ordered to

$$\sum_{i=1}^{N} \begin{pmatrix} \Delta x^{(i)} \\ \Delta y^{(i)} \\ \Delta z^{(i)} \end{pmatrix} = \sum_{i=1}^{N} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \Delta x^{(i)} \\ \Delta y^{(i)} \\ \Delta z^{(i)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{13.167}$$

This re-ordering has the same effect as forming the partial derivatives with respect to the parameters $\Delta x^{(i)}, \Delta y^{(i)}, \Delta z^{(i)}$. In vector/matrix notation, we can write

$$\sum_{i=1}^{N} \mathbf{\Delta x} = \sum_{i=1}^{N} \mathbf{B}_{NNT}^{T} \cdot \mathbf{\Delta x} = \mathbf{0}. \tag{13.168}$$

where the matrix containing the partial derivatives is called $\mathbf{B}_{NNT}^{T}$. Calling it the transposed appears odd at first glance but since we will need this and the **B** matrix together we will find this very practical later on. The index *NNT* indicates that this is only the NNT part of the complete NNR/NNT constraint matrix (Eq. 13.176).

For the NNR constraint of the coordinates, Eq. 13.165 translates to

$$\sum_{i=1}^{N} (\mathbf{r}^{(i)} \times \mathbf{\Delta x}^{(i)}) = \sum_{i=1}^{N} \left(\begin{pmatrix} x^{(i)} \\ y^{(i)} \\ z^{(i)} \end{pmatrix} \times \begin{pmatrix} \Delta x^{(i)} \\ \Delta y^{(i)} \\ \Delta z^{(i)} \end{pmatrix} \right) = \sum_{i=1}^{N} \begin{pmatrix} y^{(i)} \cdot \Delta z^{(i)} - z^{(i)} \cdot \Delta y^{(i)} \\ z^{(i)} \cdot \Delta x^{(i)} - x^{(i)} \cdot \Delta z^{(i)} \\ x^{(i)} \cdot \Delta y^{(i)} - y^{(i)} \cdot \Delta x^{(i)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{13.169}$$

If we re-order the argument of the sum into a 3×3 matrix by

$$\begin{pmatrix} y^{(i)} \cdot \Delta z^{(i)} - z^{(i)} \cdot \Delta y^{(i)} \\ z^{(i)} \cdot \Delta x^{(i)} - x^{(i)} \cdot \Delta z^{(i)} \\ x^{(i)} \cdot \Delta y^{(i)} - y^{(i)} \cdot \Delta x^{(i)} \end{pmatrix} = \begin{pmatrix} 0 & y^{(i)} \cdot \Delta z^{(i)} & -z^{(i)} \cdot \Delta y^{(i)} \\ -x^{(i)} \cdot \Delta z^{(i)} & 0 & z^{(i)} \cdot \Delta x^{(i)} \\ x^{(i)} \cdot \Delta y^{(i)} & -y^{(i)} \cdot \Delta x^{(i)} & 0 \end{pmatrix} \tag{13.170}$$

and again separating the coordinate components into a 3-line vector and the coefficients into a 3×3 matrix, which is in fact again as if we had formed the partial derivatives, we find

$$\begin{pmatrix} y^{(i)} \cdot \Delta z^{(i)} - z^{(i)} \cdot \Delta y^{(i)} \\ z^{(i)} \cdot \Delta x^{(i)} - x^{(i)} \cdot \Delta z^{(i)} \\ x^{(i)} \cdot \Delta y^{(i)} - y^{(i)} \cdot \Delta x^{(i)} \end{pmatrix} = \begin{pmatrix} 0 & -z^{(i)} & y^{(i)} \\ z^{(i)} & 0 & -x^{(i)} \\ -y^{(i)} & x^{(i)} & 0 \end{pmatrix} \cdot \begin{pmatrix} \Delta x^{(i)} \\ \Delta y^{(i)} \\ \Delta z^{(i)} \end{pmatrix}. \tag{13.171}$$

with the 3×3 matrix containing the partial derivatives. As above

$$\sum_{i=1}^{N} \begin{pmatrix} 0 & -\tilde{z}^{(i)} & \tilde{y}^{(i)} \\ \tilde{z}^{(i)} & 0 & -\tilde{x}^{(i)} \\ -\tilde{y}^{(i)} & \tilde{x}^{(i)} & 0 \end{pmatrix} \cdot \begin{pmatrix} \Delta x^{(i)} \\ \Delta y^{(i)} \\ \Delta z^{(i)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{13.172}$$

in vector notation represents

$$\sum_{i=1}^{N} \mathbf{B}_{NNR}^{T} \cdot \mathbf{\Delta x} = \mathbf{0}. \tag{13.173}$$

For the final NNR/NNT datum model for the coordinates, we can add the translation constraint and the rotation constraint (Eqs. 13.167 and 13.172) to a composite constraint with the sum over all telescopes:

$$\sum_{i=1}^{N} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -\tilde{z}^{(i)} & \tilde{y}^{(i)} \\ \tilde{z}^{(i)} & 0 & -\tilde{x}^{(i)} \\ -\tilde{y}^{(i)} & \tilde{x}^{(i)} & 0 \end{pmatrix} \cdot \begin{pmatrix} \Delta x^{(i)} \\ \Delta y^{(i)} \\ \Delta z^{(i)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{13.174}$$

Again, this is

$$\sum_{i=1}^{N} \mathbf{B}_{coord}^{T} \cdot \mathbf{\Delta x} = \mathbf{0}. \tag{13.175}$$

with the index *coord* indicating that it refers to the coordinates.

As a brief sideline, it should be mentioned that, except of the scale m elements, the transposed of the matrix in Eq. 13.174 corresponds to the **H** matrix in the datum definition by rendering with Helmert parameters (Eq. 13.206). For this reason, the application of the NNR/NNT constraints to the datum-free normal matrix has the same effect as rendering the VLBI normal equation system with Helmert parameters for this purpose (Sec. 13.8.6).

The fact, that Eq. 13.174 is for one telescope only, requires that we repeat this for all other datum telescopes as

$$\mathbf{B}_{coord}^{T} = \begin{array}{c} \begin{array}{ccccccccccc} \Delta x^{(1)} & \Delta y^{(1)} & \Delta z^{(1)} & \dots & \Delta x^{(2)} & \Delta y^{(2)} & \Delta z^{(2)} & \dots & \Delta x^{(N)} & \Delta y^{(N)} & \Delta z^{(N)} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 1 \\ 0 & -\tilde{z}^{(1)} & \tilde{y}^{(1)} & \dots & 0 & -\tilde{z}^{(2)} & \tilde{y}^{(2)} & \dots & 0 & -\tilde{z}^{(N)} & \tilde{y}^{(N)} \\ \tilde{z}^{(1)} & 0 & -\tilde{x}^{(1)} & \dots & \tilde{z}^{(2)} & 0 & -\tilde{x}^{(2)} & \dots & \tilde{z}^{(N)} & 0 & -\tilde{x}^{(N)} \\ -\tilde{y}^{(1)} & \tilde{x}^{(1)} & 0 & \dots & -\tilde{y}^{(2)} & \tilde{x}^{(2)} & 0 & \dots & -\tilde{y}^{(N)} & \tilde{x}^{(N)} & 0 \end{pmatrix} \end{array} \tag{13.176}$$

In block matrix notation we can write $\mathbf{B}_{coord}$ for all parameters including those for the coordinates of the datum telescopes (i) as

$$\mathbf{B}_{coord}^{T} = \begin{pmatrix} \mathbf{B}_{(1)}^{T} & \mathbf{B}_{(2)}^{T} & \dots & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{B}_{(N)}^{T} \end{pmatrix} \tag{13.177}$$

where the **0** are located in columns of parameters which are unaffected by the NNR/NNT rule, such as telescopes with uncertain coordinates or other parameters of the VLBI solution. This means that all columns of the **B** matrix for the telescope coordinates, for which the NNR/NNT datum constraints are supposed to be applied, are populated with the respective numbers of the 6-line matrix in Eq. 13.174 while all other elements are zero. The dimension of this matrix is $6 \times (3N + P)$ with P being all other non-datum parameters. How we use $\mathbf{B}^T_{coord}$ (and $\mathbf{B}_{coord}$ will be explained after we have looked at the NNR/NNT condition for the velocities.

When dealing with ensembles of sessions, we have to take into account that telescopes are moving due to tectonic activities, most prominently continental drift. Since this, in most cases, can be considered as a linear phenomenon, we also have to apply a datum for the telescope's velocities. Otherwise, the network could produce a net rotation which would corrupt the estimated Earth orientation parameters. In fact, this was the initial reason for introducing the NNR constraints (Sec. 13.8.3). In addition, the whole network could virtually leave the Earth's surface because of a net translation in the velocities.

The NNR and NNT constraints of the velocities **v** are formulated in a similar way as for the coordinate differences $\Delta\mathbf{x}$ as above. We assume that the velocities of the VLBI frame are close to those of the conventional or target reference frame, i.e., that $\Delta\mathbf{v} = \mathbf{v} - \tilde{\mathbf{v}}$ are small. Then

$$\sum_{i=1}^{N} \Delta\mathbf{v}^{(i)} = \mathbf{0} \tag{13.178}$$

is the No-net-translation constraint for the velocities. In vector component writing, these constraints are

$$\sum_{i=1}^{N} \Delta\mathbf{v}^{(i)} = \sum_{i=1}^{N} \begin{pmatrix} \Delta vx^{(i)} \\ \Delta vy^{(i)} \\ \Delta vz^{(i)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{13.179}$$

and re-ordered again to form the partial derivatives at the correct locations to

$$\begin{pmatrix} vx^{(i)} - \widetilde{vx}^{(i)} \\ vy^{(i)} - \widetilde{vy}^{(i)} \\ vz^{(i)} - \widetilde{vz}^{(i)} \end{pmatrix} = \begin{pmatrix} \Delta vx^{(i)} \\ \Delta vy^{(i)} \\ \Delta vz^{(i)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \Delta vx^{(i)} \\ \Delta vy^{(i)} \\ \Delta vz^{(i)} \end{pmatrix} \tag{13.180}$$

For the No-net-rotation constraint of the velocities, we cannot just apply the NNR constraint as we have done for the coordinates, but here we need to introduce a special relation for the constraint, namely that

$$\sum_{i=1}^{N} (\mathbf{r}^{(i)} \times \mathbf{v}^{(i)}) - \sum_{i=1}^{N} (\tilde{\mathbf{r}}^{(i)} \times \tilde{\mathbf{v}}^{(i)}) = \mathbf{0} \tag{13.181}$$

with N datum telescopes (Angermann et al., 2004). With $\mathbf{r}^{(i)} = \tilde{\mathbf{r}}^{(i)} + \Delta\mathbf{x}$ and $\mathbf{v}^{(i)} = \tilde{\mathbf{v}}^{(i)} + \Delta\mathbf{v}$, Eq. 13.181 can be rewritten to

$$\sum_{i=1}^{N}\big((\tilde{\mathbf{r}}^{(i)}\times\tilde{\mathbf{v}}^{(i)})+(\mathbf{\Delta x^{(i)}}\times\tilde{\mathbf{v}}^{(i)})+(\tilde{\mathbf{r}}^{(i)}\times\mathbf{\Delta v}^{(i)})+(\mathbf{\Delta r}^{(i)}\times\mathbf{\Delta v}^{(i)})\big)-\sum_{i=1}^{N}\big(\tilde{\mathbf{r}}^{(i)}\times\tilde{\mathbf{v}}^{(i)}\big)=\mathbf{0}. \quad (13.182)$$

With the last summand within the first sum to be neglected because it is small, the NNR constraint for the velocities finally reads

$$\sum_{i=1}^{N}\big((\mathbf{\Delta x^{(i)}}\times\tilde{\mathbf{v}}^{(i)})+(\tilde{\mathbf{r}}^{(i)}\times\mathbf{\Delta v}^{(i)})\big)=\mathbf{0}. \quad (13.183)$$

Then in vector component writing again, these constraints are

$$\sum_{i=1}^{N}\Big(\begin{pmatrix}\Delta x^{(i)}\\ \Delta y^{(i)}\\ \Delta z^{(i)}\end{pmatrix}\times\begin{pmatrix}\widetilde{vx}^{(i)}\\ \widetilde{vy}^{(i)}\\ \widetilde{vz}^{(i)}\end{pmatrix}+\begin{pmatrix}\tilde{x}^{(i)}\\ \tilde{y}^{(i)}\\ \tilde{z}^{(i)}\end{pmatrix}\times\begin{pmatrix}\Delta vx^{(i)}\\ \Delta vy^{(i)}\\ \Delta vz^{(i)}\end{pmatrix}\Big)=\begin{pmatrix}0\\0\\0\end{pmatrix} \quad (13.184)$$

and with the cross products resolved, this is

$$\sum_{i=1}^{N}\left(\begin{pmatrix}\Delta y^{(i)}\cdot\widetilde{vz}^{(i)}-\Delta z^{(i)}\cdot\tilde{vy}^{(i)}\\ \Delta z^{(i)}\cdot\widetilde{vx}^{(i)}-\Delta x^{(i)}\cdot\tilde{vz}^{(i)}\\ \Delta x^{(i)}\cdot\widetilde{vy}^{(i)}-\Delta y^{(i)}\cdot\tilde{vx}^{(i)}\end{pmatrix}+\begin{pmatrix}\tilde{y}^{(i)}\cdot\Delta vz^{(i)}-\tilde{z}^{(i)}\cdot\Delta vy^{(i)}\\ \tilde{z}^{(i)}\cdot\Delta vx^{(i)}-\tilde{x}^{(i)}\cdot\Delta vz^{(i)}\\ \tilde{x}^{(i)}\cdot\Delta vy^{(i)}-\tilde{y}^{(i)}\cdot\Delta vx^{(i)}\end{pmatrix}\right)=\begin{pmatrix}0\\0\\0\end{pmatrix} \quad (13.185)$$

As for the coordinates, we can again re-order the matrix expressions to end up with the partial derivatives with respect to the unknown Δ's

$$\sum_{i=1}^{N}\Big(\begin{pmatrix}0 & \widetilde{vz}^{(i)} & -\widetilde{vy}^{(i)}\\ -\widetilde{vz}^{(i)} & 0 & \widetilde{vx}^{(i)}\\ \widetilde{vy}^{(i)} & -\widetilde{vx}^{(i)} & 0\end{pmatrix}\cdot\begin{pmatrix}\Delta x^{(i)}\\ \Delta y^{(i)}\\ \Delta z^{(i)}\end{pmatrix}+\begin{pmatrix}0 & -\tilde{z}^{(i)} & \tilde{y}^{(i)}\\ \tilde{z}^{(i)} & 0 & -\tilde{x}^{(i)}\\ -\tilde{y}^{(i)} & \tilde{x}^{(i)} & 0\end{pmatrix}\cdot\begin{pmatrix}\Delta vx^{(i)}\\ \Delta vy^{(i)}\\ \Delta vz^{(i)}\end{pmatrix}\Big)=\begin{pmatrix}0\\0\\0\end{pmatrix} \quad (13.186)$$

Then, again, we stack the respective elements of Eqs. 13.180 and 13.186, sum over all telescopes and on the left hand side find the formulation of the $\mathbf{B}^T$ matrix for the NNR constraint of the velocities:

$$\sum_{i=1}^{N}\begin{pmatrix}0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 1\\ 0 & \widetilde{vz}^{(i)} & -\widetilde{vy}^{(i)} & 0 & -\tilde{z}^{(i)} & \tilde{y}^{(i)}\\ -\widetilde{vz}^{(i)} & 0 & \widetilde{vx}^{(i)} & \tilde{z}^{(i)} & 0 & -\tilde{x}^{(i)}\\ \widetilde{vy}^{(i)} & -\widetilde{vx}^{(i)} & 0 & -\tilde{y}^{(i)} & \tilde{x}^{(i)} & 0\end{pmatrix}\cdot\begin{pmatrix}\Delta x^{(i)}\\ \Delta y^{(i)}\\ \Delta z^{(i)}\\ \Delta vx^{(i)}\\ \Delta vy^{(i)}\\ \Delta vz^{(i)}\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix} \quad (13.187)$$

which again represents

$$\sum_{i=1}^{N} \mathbf{B}_{vel}^{T} \cdot \mathbf{\Delta x} = \mathbf{0}. \tag{13.188}$$

The $\mathbf{B}^T$ matrix for telescope (1), again with the respective parameters in the headline, is then

$$\mathbf{B}_{vel(1)}^{T} = \begin{array}{c} \begin{array}{cccccc} \Delta x^{(1)} & \Delta y^{(1)} & \Delta z^{(1)} & \Delta vx^{(1)} & \Delta vy^{(1)} & \Delta vz^{(1)} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & \widetilde{vz}^{(1)} & -\widetilde{vy}^{(1)} & 0 & -\tilde{z}^{(1)} & \tilde{y}^{(1)} \\ -\widetilde{vz}^{(1)} & 0 & \widetilde{vx}^{(1)} & \tilde{z}^{(1)} & 0 & -\tilde{x}^{(1)} \\ \widetilde{vy}^{(1)} & -\widetilde{vx}^{(1)} & 0 & -\tilde{y}^{(1)} & \tilde{x}^{(1)} & 0 \end{pmatrix} \end{array} \tag{13.189}$$

The NNR constraint of the velocities requires not only elements for the velocity columns but also for the coordinate columns. However, since the velocities are small ($10^{-3} - 10^{-2}$ m/y) compared to the coordinates ($10^8 - 10^9$ m), the first three columns of each $\mathbf{B}_{(i)}^T$ are often set to zero for convenience.

Since we also apply the velocity datum with multiple telescopes, we have to resolve the sum as a sequence of matrices as in Eq. 13.176. In block matrix notation we can write this as

$$\mathbf{B}_{vel}^{T} = \begin{pmatrix} \mathbf{B}_{(1)}^{T} & \mathbf{0} & \mathbf{B}_{(2)}^{T} & \mathbf{0} & \ldots & \mathbf{0} & \ldots & \mathbf{B}_{(N)}^{T} \end{pmatrix} \tag{13.190}$$

Application of the B^T matrix

The following applies to both, $\mathbf{B}_{coord}^T$ and $\mathbf{B}_{vel}^T$, and will be generalized by $\mathbf{B}^T$. To apply the NNR/NNT datum model with the help of $\mathbf{B}^T$, we first consider the most intuitive case where we add the $\mathbf{B}^T$ matrix as extra lines to the design matrix. Then they would automatically be incorporated in the normal matrix. For the covariance matrix of the datum constraints Σ_{dd}, formal errors for these six lines have to be chosen to produce constraints with a certain weight. Either the formal errors are defined rather large (e.g., decimeters), which produces weak constraints, or would be chosen to be so small, e.g., 10^{-3} mm, that their weights become extremely large and the constraints are strong. In most applications, the formal errors are set to be identical for all datum coordinate components to keep the inner configuration of the point cloud undistorted. In other words, if some of the components or telescopes would be assigned larger or smaller formal errors, e.g., if data of certain telescopes are known to have deficits, this would distort the inner geometry of the network. For this reason, such a differentiation is applied seldom in this way but these telescopes are eliminated from the list of datum telescopes rigorously.

All this information can also be processed independently of the design and normal matrix of the observations by generating a normal matrix of the datum N_d with the dimension

$(3N + P) \times (3N + P)$:

$$\mathbf{N}_d = \mathbf{B}^T \boldsymbol{\Sigma}_{dd}^{-1} \mathbf{B} \tag{13.191}$$

$\boldsymbol{\Sigma}_{dd}^{-1}$ is the inverse of the covariance matrix of the datum constraints. The right hand side of the normal equation system of the datum is zero. Since velocities appear when time series of sessions are analysed, the process used mostly is to generate pre-reduced normal matrices of the individual sessions and stack them. In this case, it is advisable to produce a normal matrix of the datum as in Eq. 13.191 separately.

Special care has to be taken that the positions of the elements of the normal matrix of datum match those of the original normal matrix. Furthermore, for this addition to work properly, we have to scale the constraint or pseudo-observation matrix $\mathbf{B}$ for some reasons which are explained in Sec. 13.8.4. If this is guaranteed, the composition of the regularized normal matrix can be performed by adding the datum-free normal matrix N_o and the datum matrix N_d

$$\mathbf{N}_r = \mathbf{N}_o + \mathbf{N}_d = \mathbf{A}^T \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{A} + \mathbf{B}^T \boldsymbol{\Sigma}_{dd}^{-1} \mathbf{B} \tag{13.192}$$

Another way of regularizing the datum-free normal matrix $\mathbf{N}_o$ is to apply the following assumption. The solution for the transformation parameters $\boldsymbol{\Xi}$ is

$$\boldsymbol{\Xi} = (\mathbf{B}^T \boldsymbol{\Sigma}_{dd}^{-1} \mathbf{B})^{-1} \mathbf{B}^T \boldsymbol{\Sigma}_{dd}^{-1} \boldsymbol{\Delta} x \tag{13.193}$$

with $\boldsymbol{\Sigma}_{dd}^{-1}$ being a covariance matrix for the O-C vector $\boldsymbol{\Delta} x$. The unweighted matrix product

$$\mathbf{H} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \tag{13.194}$$

is regular by its composition. Thus, the minimum constraint condition may be imposed by forming the following fictitious observations

$$\mathbf{H} \cdot \boldsymbol{\Delta} x = \mathbf{0}. \tag{13.195}$$

Under this condition, the regularization of the normal matrix of VLBI observations $\mathbf{N}_o$ can also be performed with

$$\mathbf{N}_r = \mathbf{N}_o + \mathbf{N}_d = \mathbf{A}^T \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{A} + \mathbf{H} \boldsymbol{\Sigma}_{dd}^{-1} \mathbf{H}^T \tag{13.196}$$

However, as a drawback of this approach, it cannot be ignored that the necessary inversion needed for generating $\mathbf{H}$ is computationally costly, especially for many datum telescopes. Because of the regularity of the matrix $\mathbf{B}^T \mathbf{B}$, we can also use $\mathbf{H} = \mathbf{B}^T$ which then results in the same expression as in Eq. 13.192.

13.8.4. Scaling of the constraint matrix B

The theoretical need for a scaling of the constraint matrix **B** arises from the fact that the use of the straight forward geometric formulation, such as the NNR/NNT logic, is an alternative to apply a singular value decomposition of the rank defect normal matrix $\mathbf{N_o}$. Without going into too much detail, it can be said that the datum-free VLBI normal equation system with rank defect six has six Eigenvalues of $\lambda_i = 0$. The Eigenvectors belonging to these Eigenvalues can be separated into a matrix **G** (not shown here), which, in the end, belongs to the same Eigenvector system as **B** (Niemeier, 2002).

For computational purposes, this fact can be exploited by just normalizing each column of **B** to one. $\mathbf{B}_{coord}$ is the transposed of $\mathbf{B}^T_{coord}$ in Eq. 13.176:

$$\mathbf{B}_{coord} = \begin{pmatrix} 1 & 0 & 0 & 0 & \tilde{z}^{(1)} & -\tilde{y}^{(1)} \\ 0 & 1 & 0 & -\tilde{z}^{(1)} & 0 & \tilde{x}^{(1)} \\ 0 & 0 & 1 & \tilde{y}^{(1)} & -\tilde{x}^{(1)} & 0 \\ 1 & 0 & 0 & 0 & \tilde{z}^{(2)} & -\tilde{y}^{(2)} \\ 0 & 1 & 0 & -\tilde{z}^{(2)} & 0 & \tilde{x}^{(2)} \\ 0 & 0 & 1 & \tilde{y}^{(2)} & -\tilde{x}^{(2)} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & \tilde{z}^{(N)} & -\tilde{y}^{(Ni)} \\ 0 & 1 & 0 & -\tilde{z}^{(N)} & 0 & \tilde{x}^{(N)} \\ 0 & 0 & 1 & \tilde{y}^{(N)} & -\tilde{x}^{(N)} & 0 \end{pmatrix} \tag{13.197}$$

For the first three columns the normalization is achieved by multiplying each element with $1/\sqrt{N}$ where N is the number of datum telescopes. For the other columns, a normalization factor f_n^{coord} is introduced which is computed as

$$f_n^{coord} = \sqrt{\sum_{i=1}^{N} [\tilde{x}^2_{(i)} + \tilde{y}^2_{(i)} + \tilde{z}^2_{(i)}]} \tag{13.198}$$

The normalized coordinates are then $\hat{x}^{(i)} = \tilde{x}^{(i)}/f_n^{coord}$, $\hat{y}^{(i)} = \tilde{y}^{(i)}/f_n^{coord}$, and $\hat{z}^{(i)} =$

$\tilde{z}^{(i)}/f_n^{coord}$ and the normalized **B** matrix then reads

$$\mathbf{B}_{coord}^{norm} = \begin{pmatrix} 1/\sqrt{N} & 0 & 0 & 0 & \hat{z}^{(1)} & -\hat{y}^{(1)} \\ 0 & 1/\sqrt{N} & 0 & -\hat{z}^{(1)} & 0 & \hat{x}^{(1)} \\ 0 & 0 & 1/\sqrt{N} & \hat{y}^{(1)} & -\hat{x}^{(1)} & 0 \\ 1/\sqrt{N} & 0 & 0 & 0 & \hat{z}^{(2)} & -\hat{y}^{(2)} \\ 0 & 1/\sqrt{N} & 0 & -\hat{z}^{(2)} & 0 & \hat{x}^{(2)} \\ 0 & 0 & 1/\sqrt{N} & \hat{y}^{(2)} & -\hat{x}^{(2)} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1/\sqrt{N} & 0 & 0 & 0 & \hat{z}^{(N)} & -\hat{y}^{(Ni)} \\ 0 & 1/\sqrt{N} & 0 & -\hat{z}^{(N)} & 0 & \hat{x}^{(N)} \\ 0 & 0 & 1/\sqrt{N} & \hat{y}^{(N)} & -\hat{x}^{(N)} & 0 \end{pmatrix} \tag{13.199}$$

Turning to the velocities, the **B** matrix is derive by transposing Eq. 13.189

$$\mathbf{B}_{vel} = \begin{pmatrix} 0 & 0 & 0 & 0 & -\widetilde{vz}^{(1)} & \widetilde{vy}^{(1)} \\ 0 & 0 & 0 & \widetilde{vz}^{(1)} & 0 & -\widetilde{vx}^{(1)} \\ 0 & 0 & 0 & -\widetilde{vy}^{(1)} & \widetilde{vx}^{(1)} & 0 \\ 1 & 0 & 0 & 0 & \tilde{z}^{(1)} & -\tilde{y}^{(1)} \\ 0 & 1 & 0 & -\tilde{z}^{(i1} & 0 & \tilde{x}^{(1)} \\ 0 & 0 & 1 & \tilde{y}^{(1)} & -\tilde{x}^{(1)} & 0 \\ 0 & 0 & 0 & 0 & -\widetilde{vz}^{(2)} & \widetilde{vy}^{(2)} \\ 0 & 0 & 0 & \widetilde{vz}^{(2)} & 0 & -\widetilde{vx}^{(2)} \\ 0 & 0 & 0 & -\widetilde{vy}^{(2)} & \widetilde{vx}^{(2)} & 0 \\ 1 & 0 & 0 & 0 & \tilde{z}^{(2)} & -\tilde{y}^{(2)} \\ 0 & 1 & 0 & -\tilde{z}^{(2)} & 0 & \tilde{x}^{(2)} \\ 0 & 0 & 1 & \tilde{y}^{(2)} & -\tilde{x}^{(2)} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & -\widetilde{vz}^{(N)} & \widetilde{vy}^{(N)} \\ 0 & 0 & 0 & \widetilde{vz}^{(N)} & 0 & -\widetilde{vx}^{(N)} \\ 0 & 0 & 0 & -\widetilde{vy}^{(N)} & \widetilde{vx}^{(N)} & 0 \\ 1 & 0 & 0 & 0 & \tilde{z}^{(N)} & -\tilde{y}^{(N)} \\ 0 & 1 & 0 & -\tilde{z}^{(N)} & 0 & \tilde{x}^{(N)} \\ 0 & 0 & 1 & \tilde{y}^{(N)} & -\tilde{x}^{(N)} & 0 \end{pmatrix} \tag{13.200}$$

The normalization is performed in the same way as for the coordinates above. However, in the fourth to sixth column, we have a mix of velocity and coordinate components which differ in magnitude quite considerably. Since normalizing the velocities with the f_n^{coord} factor, which strictly speaking should also be augmented with the velocity components themselves, makes them even smaller ($\sim 10^{-9}$), these can safely be neglected. The normalized $\mathbf{B}_{vel}^{norm}$

matrix then reads

$$
\mathbf{B}_{vel}^{norm} = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
1/\sqrt{N} & 0 & 0 & 0 & \hat{z}^{(1)} & -\hat{y}^{(1)} \\
0 & 1/\sqrt{N} & 0 & -\hat{z}^{(1)} & 0 & \hat{x}^{(1)} \\
0 & 0 & 1/\sqrt{N} & \hat{y}^{(1)} & -\hat{x}^{(1)} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
1/\sqrt{N} & 0 & 0 & 0 & \hat{z}^{(2)} & -\hat{y}^{(2)} \\
0 & 1/\sqrt{N} & 0 & -\hat{z}^{(2)} & 0 & \hat{x}^{(2)} \\
0 & 0 & 1/\sqrt{N} & \hat{y}^{(2)} & -\hat{x}^{(2)} & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
1/\sqrt{N} & 0 & 0 & 0 & \hat{z}^{(N)} & -\hat{y}^{(N)} \\
0 & 1/\sqrt{N} & 0 & -\hat{z}^{(N)} & 0 & \hat{x}^{(N)} \\
0 & 0 & 1/\sqrt{N} & \hat{y}^{(N)} & -\hat{x}^{(N)} & 0
\end{pmatrix} \tag{13.201}
$$

In many software packages, the normalization is approximated by just dividing the coordinate component entries by the radius of the Earth. Strictly speaking, this is not correct because there is no scaling of the columns to "one". However, in most cases this serves its purpose to produce an invertible normal matrix. In addition, it reduces the discrepancy between the magnitudes of the translation (10^0) and rotation components (10^6) for numerical reasons. Tests have shown that even without any scaling, the NNR/NNT conditions are fulfilled to better than 10^{-5} m.

13.8.5. Helmert parameters and Helmert transformation

In the context of coordinate transformations and, derived from that, also for datum definition, the concept of Helmert with his similarity transformation is a convenient way to relate two 3D frames. In geodesy, there are two ways of coordinate transformations which determine how the Helmert transformation is formulated in terms of the signs of the rotational elements. In the first case, the coordinates of a point cloud are transformed from one datum to another maintaining the same axis definition. This is the predominant application in space geodesy and its global reference frames. In the second case, coordinates are transformed by translating and rotating them into a different set of coordinate axes, e.g., a local North, East, Vertical system. So beware of the signs when looking up the formula of the Helmert transformation in the literature.

Maintaining the axis definition, a Helmert transformation is applied to transform the coordinates of a telescope in the VLBI frame $x^{(i)}$, $y^{(i)}$, $z^{(i)}$ ($\mathbf{x}$) into a new frame with $\tilde{x}^{(i)}$, $\tilde{y}^{(i)}$, and $\tilde{z}^{(i)}$ ($\tilde{\mathbf{x}}$) being the coordinate components of the target data set (conventional reference frame) onto which the VLBI configuration will be mapped. Helmert's similarity transformation in its basic form needs 7 parameters, which are three translation parameters x_0, y_0, z_0, the scale m, and the three rotation parameters r_x, r_y, r_z around the respective axes. The Helmert transformation formula with the three rotation matrices ($\mathbf{R}$) is

$$\tilde{\mathbf{x}} = (1+m)\cdot \mathbf{R}_{r_z}\cdot \mathbf{R}_{r_y}\cdot \mathbf{R}_{r_x}\cdot \mathbf{x} + \mathbf{x}_0 \tag{13.202}$$

or in component writing

$$\begin{pmatrix} \tilde{x}^{(i)} \\ \tilde{y}^{(i)} \\ \tilde{z}^{(i)} \end{pmatrix} = (1+m)\cdot \begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{pmatrix} \cdot \begin{pmatrix} x^{(i)} \\ y^{(i)} \\ z^{(i)} \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} \tag{13.203}$$

See also App. A for a discussion of the order of rotations. The C_{ij} are products which result from the multiplication of the three rotation matrices for the three rotation angles around the respective axes (Eqs. A.2, A.3, and A.4). Beware again, that the conventions of the rotations define the signs. In space geodesy, the conventions regularly follow App. A and the matrix elements read

$$\begin{aligned}
C_{11} &= \cos r_y \cos r_z \qquad (13.204)\\
C_{12} &= -\cos r_x \sin r_z + \sin r_x \sin r_y \cos r_z\\
C_{13} &= \sin r_x \sin r_z + \cos r_x \sin r_y \cos r_z\\
C_{21} &= \cos r_y \sin r_z\\
C_{22} &= \cos r_x \cos r_z + \sin r_x \sin r_y \sin r_z\\
C_{23} &= -\sin r_x \cos r_z + \cos r_x \sin r_y \sin r_z\\
C_{31} &= -\sin r_y\\
C_{32} &= \sin r_x \cos r_y\\
C_{33} &= \cos r_x \cos r_y
\end{aligned}$$

Inserting these products into Eq. 13.203, the formulation is applicable to angles and translations of any size.

However, due to the rather complicated composition of the C_{ij} elements, the full transformation is used rather seldom. Instead, the frame of the VLBI coordinates x_i, y_i, z_i is already chosen to be close to that of a conventional frame. In this case, the translations and, in particular, the rotations become small and the rotation matrices simplify considerably because we can assume that $\cos\alpha = 1$ and $\sin\alpha = \alpha$ with the respective effect on any products. Then, the simplified Helmert transformation is

$$\begin{pmatrix} \tilde{x}^{(i)} \\ \tilde{y}^{(i)} \\ \tilde{z}^{(i)} \end{pmatrix} = (1+m) \cdot \begin{pmatrix} 1 & -r_z & r_y \\ r_z & 1 & -r_x \\ -r_y & r_x & 1 \end{pmatrix} \cdot \begin{pmatrix} x^{(i)} \\ y^{(i)} \\ z^{(i)} \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} \tag{13.205}$$

The question now is how to find the transformation parameters x_0, y_0, z_0, m, r_x, r_y, and r_z. This is done in a least-squares estimation process with redundant data and the $\mathbf{\Delta x} = \mathbf{x} - \tilde{\mathbf{x}}$ being the observables. If we perform the matrix $\cdot$ vector multiplication in Eq. 13.205 and re-order to isolate the transformation parameters, we find the three lines for telescope (i) to read

$$\begin{pmatrix} x^{(i)} \\ y^{(i)} \\ z^{(i)} \end{pmatrix} - \begin{pmatrix} \tilde{x}^{(i)} \\ \tilde{y}^{(i)} \\ \tilde{z}^{(i)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \tilde{x}^{(i)} & 0 & \tilde{z}^{(i)} & -\tilde{y}^{(i)} \\ 0 & 1 & 0 & \tilde{y}^{(i)} & -\tilde{z}^{(i)} & 0 & \tilde{x}^{(i)} \\ 0 & 0 & 1 & \tilde{z}^{(i)} & \tilde{y}^{(i)} & -\tilde{x}^{(i)} & 0 \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ y_0 \\ z_0 \\ m \\ r_x \\ r_y \\ r_z \end{pmatrix} \tag{13.206}$$

This represents the first three lines of the observation equations for the Helmert parameter estimation with the 3×7 matrix being the sub-matrix of the partial derivatives for telescope (i).

Since in many cases we also have to take into account the velocities of the telescopes, the 7-parameter Helmert transformation can be expanded to a 14-parameter similarity transformation including the velocities:

$$\begin{pmatrix} x^{(i)} \\ y^{(i)} \\ z^{(i)} \\ vx^{(i)} \\ vy^{(i)} \\ vz^{(i)} \end{pmatrix} - \begin{pmatrix} \tilde{x}^{(i)} \\ \tilde{y}^{(i)} \\ \tilde{z}^{(i)} \\ \tilde{vx}^{(i)} \\ \tilde{vy}^{(i)} \\ \tilde{vz}^{(i)} \end{pmatrix} = \begin{pmatrix} \Delta x^{(i)} \\ \Delta y^{(i)} \\ \Delta z^{(i)} \\ \Delta vx^{(i)} \\ \Delta vy^{(i)} \\ \Delta vz^{(i)} \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 0 & 0 & \tilde{x}^{(i)} & 0 & \tilde{z}^{(i)} & -\tilde{y}^{(i)} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & \tilde{y}^{(i)} & -\tilde{z}^{(i)} & 0 & \tilde{x}^{(i)} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \tilde{z}^{(i)} & \tilde{y}^{(i)} & -\tilde{x}^{(i)} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \tilde{vx}^{(i)} & 0 & \tilde{vz}^{(i)} & -\tilde{vy}^{(i)} & 1 & 0 & 0 & \tilde{x}^{(i)} & 0 & \tilde{z}^{(i)} & -\tilde{y}^{(i)} \\ 0 & 0 & 0 & \tilde{vy}^{(i)} & -\tilde{vz}^{(i)} & 0 & \tilde{vx}^{(i)} & 0 & 1 & 0 & \tilde{y}^{(i)} & -\tilde{z}^{(i)} & 0 & \tilde{x}^{(i)} \\ 0 & 0 & 0 & \tilde{vz}^{(i)} & \tilde{vy}^{(i)} & -\tilde{vx}^{(i)} & 0 & 0 & 0 & 1 & \tilde{z}^{(i)} & \tilde{y}^{(i)} & -\tilde{x}^{(i)} & 0 \end{pmatrix} \cdot$$

$$\begin{pmatrix} x_0 & y_0 & z_0 & m & r_x & r_y & r_z & \dot{x}_0 & \dot{y}_0 & \dot{z}_0 & \dot{m} & \dot{r}_x & \dot{r}_y & \dot{r}_z \end{pmatrix}^T \tag{13.207}$$

where $\dot{x}_0, \dot{y}_0, \dot{z}_0, \dot{m}, \dot{r}_x, \dot{r}_y, \dot{r}_z$ are the time derivatives of the respective transformation parameters. For more pairs, the respective number of triples or sextuplets of this type are added. In matrix notation, we can then write this as

$$\mathbf{x} - \tilde{\mathbf{x}} = \mathbf{H} \cdot \Xi. \tag{13.208}$$

$\mathbf{x} = (x_i, y_i, z_i)$ are the 3D coordinates in the original (or local) VLBI frame and $\tilde{\mathbf{x}} = (\tilde{x}_i, \tilde{y}_i, \tilde{z}_i)$ are the coordinates in the (conventional) target frame. $\mathbf{H}$ is the Jacobi matrix of the Helmert parameters which is linear already due to the pre-condition that the deltas, $\mathbf{\Delta x} = \mathbf{x} - \tilde{\mathbf{x}}$ and the rotation parameters are small. Ξ is the matrix of the seven or fourteen Helmert parameters, respectively. These parameters can be estimated through

$$\Xi = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (\mathbf{x} - \tilde{\mathbf{x}}_0). \tag{13.209}$$

With this estimation, the square sum of the residuals will be minimal under the constraint that the original polyhedron of stations is not deformed. Although six or twelve individual lines of Eqs. 13.206 or 13.207, respectively, would suffice (see 3-2-1 rule in Sec. 13.8.2), at least four (or better more, up to as many as possible) pairs of original and target telescopes with all three or six components each are employed for a reliable estimation with reduced impact of possible outliers. App. D describes the situation if only an absolute minimum of components is selected for determining the datum parameters, similar to the 3-2-1 rule above.

13.8.6. Datum definition by rendering with Helmert parameters

Eq. 13.104 of the Gauss-Markov model, which was

$$\tilde{\mathbf{x}} = (\mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{y}, \tag{13.210}$$

is the weighted least-squares (pseudo-)solution for the observations $\mathbf{y}$ without any implicit datum information causing a rank defect of six or twelve. To regularize the normal equation system of the observations $(\mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{A})$ we can render this with Helmert's translation and rotation parameters to apply a datum. This method is a strict application of the conditions

$$\mathbf{H}^T \cdot \Xi = \mathbf{w} = \mathbf{0}. \tag{13.211}$$

where $\mathbf{w}$ is the vector of contradictions. For rendering, we add the Helmert parameters to the total list of parameters and force them to be zero ($\Xi = 0$). The normal equation system in block matrix notation then is

$$\begin{pmatrix} \mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{A} & \mathbf{H} \\ \mathbf{H}^T & 0 \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{x}} \\ \mathbf{\Xi} = \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{y} \\ \mathbf{w} = \mathbf{0} \end{pmatrix}. \tag{13.212}$$

The $\mathbf{H}$ matrix is the Jacobi matrix of Eq. 13.208 and defines the datum with line triples or sextuplets for each of the N telescopes, placed at the respective lines and columns.

In general, with $\mathbf{n} = \mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{y}$ the parameter vector of the estimates is

$$\begin{pmatrix} \tilde{\mathbf{x}} \\ \mathbf{\Xi} = \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{A} & \mathbf{H} \\ \mathbf{H}^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{n} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{Q}_{\tilde{x}\tilde{x}} & \mathbf{Q}_{\tilde{x}\Xi} \\ \mathbf{Q}_{\Xi\tilde{x}} & \mathbf{Q}_{\Xi\Xi} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{n} \\ \mathbf{0} \end{pmatrix}. \tag{13.213}$$

For the determination of the coordinates, we compute

$$\tilde{\mathbf{x}} = \mathbf{Q}_{\tilde{x}\tilde{x}} \cdot \mathbf{n} \tag{13.214}$$

The addition of the Helmert parameters with pre-defined values (zeros) yields a normal matrix which is regular. In other words, this forces the adjustment to produce telescope coordinates which deviate from the reference coordinates in a way that the sum of the translations of all telescopes and the rotations of the network with respect to the reference network are forced to zero. If the scale m is deleted from the Helmert parameter list, this procedure yields results for the telescope coordinates which are equivalent to the no-net-rotation and no-net-translation (NNR/NNT) condition approach as described above. The reason is that the Jacobi matrix $\mathbf{H}$ corresponds to the $\mathbf{B}$ matrix as composed of Eqs. 13.174 and 13.189.

It should be emphasized here, that the same logic as for the NNR/NNT constraints applies in terms of singular value decomposition of the datum-free $\mathbf{N}$ matrix with rank defect six or twelve. Again, this can be by-passed in the computations by normalizing the columns of the $\mathbf{H}$ matrix as described in Sec. 13.8.4.

13.8.7. Practical issues of datum definition

First of all, the application of datum definition depends on the purpose of the solution. We can apply a datum for a single session alone and for a group of sessions as well as for the whole list of sessions available, i.e., for a global solution. In the latter two cases, the datum is applied after the ensemble of datum-free normal equation systems was stacked.

The selection of the pairs of coordinates in the original and target list depends on the quality of their coordinates, respectively the quality of the observations which led to these coordinates. It is quite understandable that telescopes with technical difficulties or those with only a short observational history compared to the rest of telescopes may not be that well suited for the estimation of session-wise Helmert parameters or NNR/NNT conditions. For this reason only a subset of telescopes is normally employed for NNR/NNT conditions or for determining the Helmert parameters as compared to using all available coordinate pairs. However, in general, the selection of datum sites should ideally have a good global coverage to avoid unwanted translations and rotations.

13.9. Celestial datum definitions

13.9.1. General and functional considerations of celestial datum definition

As for the terrestrial datum definition, almost the same general considerations apply to the celestial datum definition. Consequently, at first VLBI observations only provide relative relationships in-between the point cloud of radio source positions which are considered to be located on a unit sphere. Again, this is just the geometric *configuration* and we have to apply a datum with a datum definition to produce a meaningful celestial reference frame.

For this, conceptually there are the same six degrees of freedom as for the terrestrial frame, three translations and three rotations. However, because of the huge distances between Earth and the radio sources, and because the origin (the solar system barycenter) is implicit to all computations, we normally only consider the three rotations as applicable degrees of freedom. The scale of course is also fixed explicitly through the speed of light.

With respect to the positions of the radio sources, time dependency is not an issue yet for most of the sources. Therefore, datum definition is limited only to the three rotations as time invariant quantities.

The task of datum definition of a celestial reference frame is to define the orientation of the free network in a conventional frame to be able to produce radio source positions in a pre-defined frame. Examples of such conventional frames were described in Sec. 13.1.2. The need for this may also arise from the facts that the configuration of the polyhedron of source positions be slightly different to that of the conventional frame (because of new and better observations and analysis models) and/or because there may be new sources which shall be mapped into a conventional frame.

As in the terrestrial case, we can choose between two options for the problem of datum definition. The first option is that source positions and their covariances have already been determined in an arbitrary reference frame in a previous VLBI analysis process. We then apply a similarity transformation to rotate this frame into a conventional one, where the polyhedron of positions, respectively its configuration, is not distorted. This can be done in different ways. The second option is that we incorporate the functionality of the first option in our VLBI analysis software and apply the datum directly in the analysis. This has the advantage that we save a processing step but blank out the transformation parameters which sometimes are of interest as well.

The most common way of celestial datum definition today is the no-net-rotation (NNR) condition or constraint (Sec. 13.9.3) which considers a defined ensemble of the points in 3D space to match the conventional target frame as best as possible. Equivalent in its result is the rendering by Helmert parameters (Sec. 13.9.4). Before we turn to these methods, we look at the simplest method of datum definition which is free of distortion, the 2-1 celestial datum rule (Sec. 13.9.2).

Table 13.4: Methods of celestial datum definition.

	design/Jacobi matrix	Normal matrix		
single session datum	pseudo observations	2-1 rule	Constraint matrix $\mathbf{C}^{\mathbf{T}}\mathbf{\Sigma}_{\mathbf{cc}}^{-1}\mathbf{C}$	Helmert rendering
multi session datum	pseudo observations	2-1 rule	Constraint matrix $\mathbf{C}^{\mathbf{T}}\mathbf{\Sigma}_{\mathbf{cc}}^{-1}\mathbf{C}$	Helmert rendering

13.9.2. 2-1 celestial datum rule

As in the terrestrial case, we start with a simple approach which has in fact been in use at the early days of geodetic VLBI as well. In a VLBI solution, the celestial configuration has three degrees of freedom or a rank defect of three. The three rotational degrees are used to define the directions of the coordinate axes. This information is all embedded in the source positions. For this reason, we can just apply our definitions to some of the positions alone. Here, the clue is the 2-1 rule. We take one radio source and assign the complete pair of two position components, right ascension and declination, e.g., from the ICRF3 table, to this. Then we take a second source, ideally diametrically opposite on the celestial sphere of the first one and fix just one position component of this source to the tabulated one.

To derive the three observation equations, we first consider the basic rotation of the position vector in polar coordinates from the starting position α, δ to the target (catalog) position $\tilde{\alpha}$, $\tilde{\delta}$

$$\begin{pmatrix} \cos\tilde{\alpha}_i\cos\tilde{\delta}_i \\ \sin\tilde{\alpha}_i\cos\tilde{\delta}_i \\ \sin\tilde{\delta}_i \end{pmatrix} = \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{pmatrix} \cdot \begin{pmatrix} \cos\alpha_i\cos\delta_i \\ \sin\alpha_i\cos\delta_i \\ \sin\delta_i \end{pmatrix} \tag{13.215}$$

The C_{ij} are coefficients which result from the multiplication of three rotation matrices for the three rotation angles around the respective axes (Eqs. A.2, A.3, and A.4) containing a row of sine and cosine terms with respect to the three rotation parameters r_1, r_2, r_3 (Eq. 13.205). x_c, y_c, z_c are the respective Cartesian coordinates on a unit sphere. The formulation is applicable to angles and translations of any size.

Normally, the frame of the VLBI source positions α_i, δ_i is already chosen to be close to that of a conventional frame $\tilde{\alpha}_i, \tilde{\delta}_i$. In this case, the rotations become small and the rotation matrices simplify considerably because we can assume that $\cos\alpha = 1$ and $\sin\alpha = \alpha$. Then the right hand side of Eq. 13.215 can be written as

$$\begin{pmatrix} \cos\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin\tilde{\delta}_i \end{pmatrix} = \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = \begin{pmatrix} 1 & -r_z & r_y \\ r_z & 1 & -r_x \\ -r_y & r_x & 1 \end{pmatrix} \cdot \begin{pmatrix} \cos\alpha_i \cos\delta_i \\ \sin\alpha_i \cos\delta_i \\ \sin\delta_i \end{pmatrix} \tag{13.216}$$

For the derivation of observation equations with respect to the unknown rotation parameters, we can reformulate Eq. 13.216 to

$$\begin{pmatrix} \cos\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin\tilde{\delta}_i \end{pmatrix} = \begin{pmatrix} \cos\alpha_i \cos\delta_i \\ \sin\alpha_i \cos\delta_i \\ \sin\delta_i \end{pmatrix} + \begin{pmatrix} 0 & \sin\delta_i & -\sin\alpha_i \cos\delta_i \\ -\sin\delta_i & 0 & \cos\alpha_i \cos\delta_i \\ \sin\alpha_i & -\cos\alpha_i \cos\delta_i & 0 \end{pmatrix} \cdot \begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix} \tag{13.217}$$

or

$$\begin{pmatrix} \cos\alpha_i \cos\delta_i - \cos\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin\alpha_i \cos\delta_i - \sin\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin\delta_i - \sin\tilde{\delta}_i \end{pmatrix} = \begin{pmatrix} 0 & \sin\delta_i & -\sin\alpha_i \cos\delta_i \\ -\sin\delta_i & 0 & \cos\alpha_i \cos\delta_i \\ \sin\alpha_i & -\cos\alpha_i \cos\delta_i & 0 \end{pmatrix} \cdot \begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix} \tag{13.218}$$

Although we have three lines for three parameters, they count only for two because the Cartesian differences on the left hand side only depend on two parameters α and δ. For this reason, there is no unique solution for the rotation parameters. If the source index i in Eq. 13.218 is replaced, a fourth equation for source #2 to the effect

$$\begin{pmatrix} \sin\delta_2 - \sin\tilde{\delta}_2 \end{pmatrix} = \begin{pmatrix} \sin\alpha_2 & -\cos\alpha_2 \cos\delta_2 & 0 \end{pmatrix} \cdot \begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix} \tag{13.219}$$

needs to be added to form a solvable equation system.

Having determined the parameters r_1, r_2, and r_3, we can transform all source positions of the VLBI frame into those of the target (conventional) frame applying Eq. 13.216 in polar coordinates on a unit sphere. This leaves the configuration as it was initially. In the same way as sketched here, we can perform the fixing of the three source position components directly in the VLBI analysis program by eliminating the respective columns in the Jacobi matrix leading to an implicit datum definition.

In the early days of VLBI, this celestial datum definition was applied by selecting the position of the source 3C273B from the FK5 catalog and fixing the declination of OQ208 to the respective catalog value. Of course, the deficit of this approach was that the uncertainties of one or more of the components, which are fixed, directly affect all other components to be estimated. For this reason, this procedure is not used in operational analyses any more but was described here for didactic purposes nevertheless.

In the same category fall the following considerations as well. The datum definition as described here and also later always allow to also estimate the two components of the celestial pole offsets (CPO), dX and dY. The 2-1 rule can also be used in the way that the two components of the CPO and the right ascension of one radio source are fixed to some tabulated values (e.g., IERS C04 series) providing the three rotation definitions for the full celestial datum. Doing this, all other source positions can be estimated. In fact, the values of the IAU 1980 nutation model and the right ascension of 3C273B were used to define the celestial datum in the VLBI solutions of NASA's Crustal Dynamics Project (Ma et al., 1989).

13.9.3. Celestial datum definition with no-net-rotation conditions or constraints

In general, for celestial datum definition with no-net-rotation (NNR) conditions or constraints, the same applies as for the terrestrial datum definition. However, here we deal with angular quantities, right ascensions $\tilde{\alpha}_i$ and declinations $\tilde{\delta}_i$ of the reference or target frame, α_i and δ_i of the original VLBI positions as well as their residuals or adjustments $\Delta\alpha$ and $\Delta\delta$ in angular dimensions with $\Delta\alpha = \alpha - \tilde{\alpha}$ and $\Delta\delta = \delta - \tilde{\delta}$. Since we apply the NNR condition based on individual radio sources, it reads (Jacobs et al., 2010)

$$\sum_{i=1}^{N}(\mathbf{k}_i \times \mathbf{\Delta k}_i) = \mathbf{0} \tag{13.220}$$

with $\mathbf{k}$ containing the source position components according to Eq. 13.31 with N radio sources selected for the datum definition. The $\mathbf{\Delta k}_i$ vector needs to be written as the difference between the VLBI position to be rotated and the target position as

$$\mathbf{\Delta k}_i = \begin{pmatrix} \cos\alpha_i \cos\delta_i - \cos\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin\alpha_i \cos\delta_i - \sin\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin\delta_i - \sin\tilde{\delta}_i \end{pmatrix}. \tag{13.221}$$

Since we want to formulate the condition/constraint with the residuals $\Delta\alpha$ and $\Delta\delta$, we replace the original VLBI positions α_i and δ_i with $\alpha = \tilde{\alpha} + \Delta\alpha$ and $\delta = \tilde{\delta} + \Delta\delta$ to yield

$$\mathbf{\Delta k}_i = \begin{pmatrix} \cos(\tilde{\alpha}_i + \Delta\alpha_i)\cos(\tilde{\delta}_i + \Delta\delta_i) - \cos\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin(\tilde{\alpha}_i + \Delta\alpha_i)\cos(\tilde{\delta}_i + \Delta\delta)_i - \sin\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin(\tilde{\delta}_i + \Delta\delta_i) - \sin\tilde{\delta}_i \end{pmatrix} \tag{13.222}$$

The condition/constraint with the cross product to be computed then reads

$$\sum_{i=1}^{N}(\mathbf{k}_i \times \mathbf{\Delta k}_i) = \sum_{i=1}^{N}\left(\begin{pmatrix} \cos\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin\tilde{\delta}_i \end{pmatrix} \times \begin{pmatrix} \cos(\tilde{\alpha}_i + \Delta\alpha_i)\cos(\tilde{\delta}_i + \Delta\delta_i) - \cos\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin(\tilde{\alpha}_i + \Delta\alpha_i)\cos(\tilde{\delta}_i + \Delta\delta_i) - \sin\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin(\tilde{\delta}_i + \Delta\delta_i) - \sin\tilde{\delta}_i \end{pmatrix}\right) \tag{13.223}$$

The cross multiplication is quite a lengthy process with many sine and cosine terms. For details please see App. E. The result following Eqs. E.7, E.8, and E.9, is

$$\mathbf{k}_i \times \mathbf{\Delta k}_i = \begin{pmatrix} -\cos\tilde{\alpha}_i \sin\tilde{\delta}_i \cos\tilde{\delta}_i \Delta\alpha_i + \sin\tilde{\alpha}_i \Delta\delta_i \\ -\sin\tilde{\alpha}_i \sin\tilde{\delta}_i \cos\tilde{\delta}_i \Delta\alpha_i - \cos\tilde{\alpha}_i \Delta\delta_i \\ \cos^2\tilde{\delta}_i \Delta\alpha_i \end{pmatrix} \tag{13.224}$$

Following the same considerations as for the terrestrial datum definition, we can formulate

$$\sum_{i=1}^{N} (\mathbf{k}_i \times \mathbf{\Delta k}_i) = \sum_{i=1}^{N} \mathbf{C}^T \cdot \mathbf{\Delta k} = \mathbf{0} \tag{13.225}$$

with the $\mathbf{C}^T$ matrix containing the partial derivatives for radio source i

$$\sum_{i=1}^{N} \begin{pmatrix} -\cos\tilde{\alpha}_i \sin\tilde{\delta}_i \cos\tilde{\delta}_i & \sin\tilde{\alpha}_i \\ -\sin\tilde{\alpha}_i \sin\tilde{\delta}_i \cos\tilde{\delta}_i & -\cos\tilde{\alpha}_i \\ \cos^2\tilde{\delta}_i & 0 \end{pmatrix} \cdot \begin{pmatrix} \Delta\alpha_i \\ \Delta\delta_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{13.226}$$

Along the same lines as for the terrestrial datum definition, in block matrix writing we can define $\mathbf{C}^T$ as

$$\mathbf{C}^T = \begin{pmatrix} \mathbf{C}_1^T & \mathbf{C}_2^T & \dots & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{C}_{(N)}^T \end{pmatrix} \tag{13.227}$$

where the **0** are again located in columns of parameters which are unaffected by the NNR rule for the N datum sources.

All this information can be processed independently of the design and normal matrix of the observations by generating a normal matrix of the celestial datum N_{cd} with the dimension $(2N + P) \times (2N + P)$ (P = all non-datum parameters):

$$\mathbf{N}_{cd} = \mathbf{C}^T \mathbf{\Sigma}_{cc}^{-1} \mathbf{C} \tag{13.228}$$

$\mathbf{\Sigma}_{cc}^{-1}$ is the inverse of the covariance matrix of the constraints forming the weight matrix. This should have the same elements for all datum sources in order to not deform the configuration of the datum sources. The right hand side of the normal equation system of the datum is zero. In the same way as for the terrestrial datum, the composition of the regularized normal matrix can be performed by adding the two normal matrices

$$\mathbf{N}_r = \mathbf{N}_o + \mathbf{N}_{cd} = \mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{A} + \mathbf{C}^T \mathbf{\Sigma}_{cc}^{-1} \mathbf{C} \tag{13.229}$$

Here, the same applies for other options and possible re-weighting as described in the section of the terrestrial datum (Sec. 13.8).

13.9.4. Celestial datum definition by rendering

Finally, to complete the set of options for the celestial datum definition, we should mention the rendering concept. This could also be called truncated Helmert rendering because only the rotations are applied as constraints. In block matrix notation, the normal equation system is composed as

$$\begin{pmatrix} \mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^T & 0 \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{x}} \\ \mathbf{\Xi} = \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{y} \\ \mathbf{w} = \mathbf{0} \end{pmatrix}. \tag{13.230}$$

with $\mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{A}$ being the celestial datum free normal matrix, $\mathbf{C}$ being the partial derivatives with respect to α and δ as in Sec. 13.9.3, $\tilde{x}$ contains the unknown standard VLBI parameters, Ξ consists of the rotations α and δ forced to be zero. On the right hand side, $\mathbf{A}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{y}$ contains the observed minus computed elements $\mathbf{y}$ of the VLBI delay observations. $\mathbf{w}$ is the vector of source position residuals or contradictions which has to be forced to zero as well. As for the terrestrial datum definition, rendering applies conditions which have the same effect as very strict constraints.

14. Combination of results

In many publications (e.g., Böckmann et al. 2010), it has been shown that the combination of multiple solutions produces superior results compared to solutions of individual analysis centers (ACs). Within the IVS, the combination is performed with input from a number of IVS ACs. The exact number varies and depends on the purpose of the combination. For operational combinations producing predominantly Earth orientation parameters for low-latency dissemination, only five to seven ACs tend to meet the timeliness guidelines. In general, the number for the computation of the IVS input to ITRF is larger. For ITRF2020 (Hellmers et al., 2022) it was eleven ACs with seven different analysis software packages.

Although a rough combination may be possible with averaging all EOP results or do a combination of results with full covariance information, the IVS at a very early stage of its existence decided to combine the output of its ACs on the basis of pre-reduced datum-free normal equation systems (Nothnagel, 2002). This means that the solutions are not computed to the last least squares inversion step but that the original normal equation system is split into a part, which contains only the parameters to be combined ($\mathbf{N_{11}}$), such as the telescope coordinates and EOP, and a part ($\mathbf{N_{22}}$) which contains the parameters inherent to the solution setup of the AC, such as clock and atmosphere parameters. The benefit is that the full geometric solution configuration is available in the combination process allowing rigorous epoch transformations and homogenization of the reference frames to be used (Vennebusch et al., 2007). In other words, this guarantees that the same terrestrial and in some cases also the celestial reference frame is applied without the need for prior organizational complications.

From a conceptual point of view, the whole adjustment process can be considered as a stepwise least squares adjustment. The first step is the preparation of a pre-reduced normal equation system for each session. This can be used to complete the inversion in a second reduction step and provide solutions for all parameters of the sessions. Forking to the combination process, the pre-reduced systems can be combined with systems of other solutions through the theorem of addition of normal equation systems (Koch, 1999). Helmert (1872) had developed this concept to link triangulation networks in the absence of electronic computing power. After the addition is carried out, which is also called stacking, the combined normal equation system is reduced further to solve for all parameters as is done for a single session.

14.1. Preparation of pre-reduced normal equation systems

Initially, the structure of the normal equation system of each IVS analysis center consists of the normal matrix (Eq. 13.101) and the right hand side (Eq. 13.102). The normal equation system is free of datum and should have a rank defect of six for the terrestrial reference frame and an additional three for the celestial reference frame if radio source position parameters are included (see also Sec. 13.8.1). The system can be re-ordered as Eq. 14.1, where $\mathbf{x_1}$ is the vector of parameters to be combined later such as telescope coordinates, EOP, and radio source positions (if included). $\mathbf{x_2}$ is the vector of local parameters, such as for the clocks and atmospheres. $\mathbf{b_1}$ and $\mathbf{b_2}$ are the respective right hand sides.

$$\begin{pmatrix} \mathbf{N_{11}} & \mathbf{N_{12}} \\ \mathbf{N_{21}} & \mathbf{N_{22}} \end{pmatrix} \begin{pmatrix} \mathbf{x_1} \\ \mathbf{x_2} \end{pmatrix} = \begin{pmatrix} \mathbf{b_1} \\ \mathbf{b_2} \end{pmatrix} \tag{14.1}$$

$\mathbf{x_2}$ are the parameters which are to be pre-eliminated by rigorous reduction from Eq. 14.1 by multiplying the second line by $-\mathbf{N_{12}N_{22}^{-1}}$ from the left (Brockmann, 1997) which yields

$$\begin{pmatrix} \mathbf{N_{11}} & \mathbf{N_{12}} \\ -\mathbf{N_{12}}\,\mathbf{N_{22}^{-1}}\,\mathbf{N_{21}} & -\mathbf{N_{12}} \end{pmatrix} \begin{pmatrix} \mathbf{x_1} \\ \mathbf{x_2} \end{pmatrix} = \begin{pmatrix} \mathbf{b_1} \\ -\mathbf{N_{12}}\,\mathbf{N_{22}^{-1}}\,\mathbf{b_2} \end{pmatrix}. \tag{14.2}$$

By forming the matrix multiplications and adding the resulting two equation systems we find

$$(\mathbf{N_{11}} - \mathbf{N_{12}}\,\mathbf{N_{22}^{-1}}\,\mathbf{N_{21}})\,\mathbf{x_1} = \mathbf{b_1} - \mathbf{N_{12}}\,\mathbf{N_{22}^{-1}}\,\mathbf{b_1}. \tag{14.3}$$

The expression in parenthesis is the new normal matrix equaling the respective right hand side also reduced by the local parameters set up individually by each analysis center. For the pre-reduced normal equation system, we then have

$$\hat{\mathbf{N}}_{\mathbf{11}}\,\mathbf{x_1} = \hat{\mathbf{b}}_{\mathbf{1}}. \tag{14.4}$$

For carrying over some information on the quality of the pre-reduction to the combination step and for later assigning relative weights to the individual solutions, we also have to compute the "weighted square sum of O-C" of the reduced normal equation system. For the original system, this was $\mathbf{y}^T\,\mathbf{\Sigma}_{yy}^{-1}\,\mathbf{y}$ with the original observations y_i and the inverse of the covariance matrix $\mathbf{\Sigma}_{yy}$ which is the weight matrix $\mathbf{P}$. To take into account the influence of the pre-reduced parameters $\mathbf{x}_2$ on the "weighted square sum of O-C" of the reduced normal equation system, the product of the squared right hand side vector and the inverse of $\mathbf{N_{22}}$, which is $\mathbf{b}_2^T\,\mathbf{N_{22}}^{-1}\,\mathbf{b}_2$, is subtracted to yield the reduced value $(\mathbf{y}^T\,\mathbf{\Sigma}_{yy}^{-1}\,\mathbf{y})_{red}$ (Brockmann, 1997)

$$(\mathbf{y}^T\,\mathbf{\Sigma}_{yy}^{-1}\,\mathbf{y})_{red} = \mathbf{y}^T\,\mathbf{\Sigma}_{yy}^{-1}\,\mathbf{y} - (\mathbf{b}_2^T\,\mathbf{N_{22}}^{-1}\,\mathbf{b}_2) \tag{14.5}$$

An issue, which pops up in the combination of VLBI solutions, and here in particular in the computation of Eq. 14.5, is the fact that often large clock offsets dominate the size of the original $(\mathbf{y}^T\,\mathbf{\Sigma}_{yy}^{-1}\,\mathbf{y})$ quantity. To reduce the magnitudes of the $\mathbf{y}$ entries in the whole adjustment process mainly for numerical reasons, many IVS Analysis Centers perform a subtraction of the dominant clock effect. This can either be done by subtracting a single clock offset per remote telescope or by subtracting a pre-determined clock polynomial.

The question is, how this affects the computation of the $(\mathbf{y}^T\,\mathbf{\Sigma}_{yy}^{-1}\,\mathbf{y})_{red}$ in Eq. 14.5. Let us assume that the correlator staff detected a large clock offset during fringe search, for example, a 1 s glitch which happened to old Mark III formatters every now and then. If this glitch had been corrected at the correlator but the information was not carried over to the observed delays, then the Level-2 data analysts would not know about it. The same is of course valid for any size clock offset, which may have been eliminated at the correlator but also in a first Level-2 analysis step, e.g., in a *First Solution* of VieVS. If $\mathbf{y}_{icm}$ is the time offset of the initial clock offset, the $\mathbf{b}_2$ $(= \mathbf{A}^T\,\mathbf{\Sigma}_{yy}^{-1}\,\mathbf{y})$ vector is only computed with the reduced observations $\mathbf{y}' = \mathbf{y} - \mathbf{y_{icm}}$. In most cases, the $\mathbf{y}_{icm}$ is not re-introduced for the computation of $(\mathbf{y}^T\,\mathbf{\Sigma}_{yy}^{-1}\,\mathbf{y})_{red}$ in Eq. 14.5 and the numerical value is, thus, incorrect.

Currently, most IVS analysis centers abbreviate the observation vector $\mathbf{y}$ by subtracting a rudimentary clock polynomial to reduce the magnitude of the O-C quantities. However, in the Calc/Solve analysis package, this is not done and apparently, the subtraction of the pre-reduction step in Eq. 14.5 may not be performed either. Consequently, the WEIGHTED SQUARE SUM OF O-C may still contain rather large clock offsets. For this reason, the entries of the individual analysis centers show two distinct groups, the one on the order of 10^3 with large clock offsets pre-eliminated and the other with original observations on the order of 10^{14}. To end up with consistent meaningful "weighted square sums of O-C" for all solutions for the combination process, the author recommends that all IVS Analysis Centers only subtract a single clock offset $\mathbf{y_{icm}}$ per remote telescope, which ideally should be agreed upon between all ACs. However, this deems rather impractical and the $(\mathbf{y}^T\,\mathbf{\Sigma}_{yy}^{-1}\,\mathbf{y})_{red}$ values aggregated over all ACs need to be tagged with a label "inconsistent".

It should be emphasized that $(\mathbf{y}^T\,\mathbf{\Sigma}_{yy}^{-1}\,\mathbf{y})_{red}$ is the only additional statistical quantity which we can compute in the absence of knowing the parameters $\tilde{\mathbf{x}}_1$. We could only find these if we carry out a full inversion of the original normal equation system but these are of course not the parameters which we would find after combination.

Finally, the "weighted square sum of O-C" is often also called *ell-t-p-ell* since the character ℓ is used for the original abbreviated observations (i.e., reduced for modelled observations) in historical German statistics text books. With $\mathbf{P} = \mathbf{\Sigma}_{yy}^{-1}$, we can also write

$$(\boldsymbol{\ell}^T \mathbf{P} \boldsymbol{\ell})_{red} = (\mathbf{y}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{y})_{red}. \tag{14.6}$$

14.2. Interim storage in SINEX files

To facilitate the exchange of results and interim information, the IERS invented conventional storage in SINEX format (Solution INdependent EXchange format). Initially, the SINEX format was developed to report solutions with full covariance information[26]. Most of the entries in the SINEX definition were therefore introduced for this purpose. The option of reporting normal equation systems and the respective auxiliary information was agreed upon later. For this reason, care has to be taken to what entries can actually be reported in the SINEX submissions because sometimes they are mutually exclusive.

As stated before, the IVS had decided to provide interim results as datum-free pre-reduced normal equation systems. The SINEX files for each session have to contain the following blocks:

- FILE/REFERENCE
- FILE/COMMENT
- SITE/ID
- SOURCE/ID
- SITE/ANTENNA
- SITE/ECCENTRICITY
- SOLUTION/EPOCHS
- NUTATION/DATA
- PRECESSION/DATA
- SOLUTION/STATISTICS
- SOLUTION/APRIORI
- SOLUTION/NORMAL_EQUATION_MATRIX
- SOLUTION/NORMAL_EQUATION_VECTOR
- SOLUTION/STATISTICS

The content and format of these entries are described accurately in the reference document https://www.iers.org/IERS/EN/Organization/AnalysisCoordinator/SinexFormat/sinex.html and are basically straight forward. The only exception is the SOLUTION/STATISTICS block. For VLBI, the following entries are of interest but again some are restricted to full solutions and some are valid for pre-reduced normal equation systems:

[26] https://www.iers.org/SharedDocs/Publikationen/EN/IERS/Documents/ac/sinex/sinex_v202_pdf.pdf

- NUMBER OF OBSERVATIONS = # of observations used in the adjustment
  This entry is valid for full solutions and pre-reduced normal equation systems. The statement in the reference document reads "The 'NUMBER OF OBSERVATIONS' should represent only the number of 'real' observations.". Although the number of constraint equations also belongs to the group of observations, they are not accounted for here because strictly speaking these should be listed elsewhere. Since they only implicitly appear in the NUMBER OF DEGREES OF FREEDOM (see below), they are often included here.

- NUMBER OF UNKNOWNS = # of unknowns solved in the adjustment
  This entry is valid for full solutions and pre-reduced normal equation systems. In both cases it refers to the full list of parameters. In the case of pre-reduced normal equation systems, the number of pre-reduced parameters can be computed by subtracting the dimension (number of lines or columns) of the normal equation system from the total number of unknowns.

- NUMBER OF DEGREES OF FREEDOM (df)
  In the current version of the SINEX format (2.02), this entry is the only instance of the pre-reduced normal equation system storage where the number of constraints is stored and this only implicitly "(19) dof = n_obs + n_constr – n_unk". Otherwise the entry would be redundant if NUMBER OF OBSERVATIONS and NUMBER OF UNKNOWNS are stated correctly.

- WEIGHTED SQUARE SUM OF O-C = $\boldsymbol{\ell}^T \mathbf{P} \boldsymbol{\ell}_{red}$
  This entry is valid for full solutions and pre-reduced normal equation systems. If the computation of the dimension used in the weight matrix is consistent with the dimensions of the observations, the entry is unitless (Sec. 13.7.1). Care should be taken of how any pre-eliminated contributions to the observations have been handled (see Sec. 14.1).

- SQUARE SUM OF RESIDUALS (VTPV) = Sum of squares of residuals
  This entry can only be filled for full solutions because all parameters need to be available for the computation of the residuals ($\mathbf{v} = \mathbf{A}\tilde{\mathbf{x}} - \mathbf{y}$). In the case of step-wise adjustments or combinations, we initially only have the weighted sum squared of O-C $\mathbf{y}^T \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{y}$ with the original observations y_i and the inverse of the covariance matrix $\boldsymbol{\Sigma}_{yy}$ (= weight matrix $\mathbf{P}$). If we know all parameters $\tilde{\mathbf{x}}$ we subtract the weighted sum squared of the computed observations $\mathbf{y}^T \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{A} \tilde{\mathbf{x}}$ and compute the sum of weighted squares of residuals with

$$\mathbf{v}^T \mathbf{P} \mathbf{v} = \mathbf{y}^T \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{y} - \mathbf{y}^T \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{A} \tilde{\mathbf{x}}. \tag{14.7}$$

In some publications, also the following may be in use

$$\chi^2 = \tilde{\Omega} = \mathbf{v}^T \mathbf{P} \mathbf{v}. \tag{14.8}$$

The expression χ^2 is mentioned in the SINEX format document and should not be confused with the same expression in Sec. 13.7.1 for testing the balance between the variance of unit weight a priori and that a posteriori. There also is a χ^2 statistical test related to the goodness of fit.

For completeness, it should be mentioned that we may also compute $\mathbf{v}^T \mathbf{P} \mathbf{v}$ after the combination has been performed and the combined parameters $\tilde{\mathbf{x}}_1$ are available. We can then also write (Brockmann, 1997),

$$\begin{aligned} \mathbf{v}^T \mathbf{P} \mathbf{v} &= \mathbf{y}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{y} - \tilde{\mathbf{x}}_1^T \mathbf{b}_1 - \mathbf{b}_2^T \mathbf{N}_{22}^{-1} \mathbf{b}_2 \\ &= \mathbf{y}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{y}_{red} - \tilde{\mathbf{x}}_1^T \mathbf{b}_1 \end{aligned} \tag{14.9}$$

Before this goes unnoticed, we should have a look at the dimensions of the vectors and matrices producing the scalar $\mathbf{v}^T \mathbf{P} \mathbf{v}$. To abbreviate the derivation, we just start with $\mathbf{x}_1^T \mathbf{b}_1 = \mathbf{y}^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{A} \tilde{\mathbf{x}}_1$. The dimensions of the components are $(1xu_1) \cdot (u_1 x1) = (1xn) \cdot (nxn) \cdot (nxu_1) \cdot (u_1 x1)$. This means that we only use those columns of the original $\mathbf{A}$ matrix which belong to the u_1 parameters. Otherwise the matrix and vector dimensions would not fit.

- VARIANCE FACTOR
  Again, this entry can only be filled for full solutions. The variance factor of unit weight (a posteriori) comes along in different names and is computed with

$$\tilde{\sigma}_0^2 = \chi^2/df = \frac{\mathbf{v}^T \mathbf{P} \mathbf{v}}{(n+n_c)-(u_1+u_2)} \tag{14.10}$$

  where n is the number of VLBI observations, n_c the number of constraint equations introduced to stabilize the estimation of clock and atmosphere parameters. u_1 and u_2 are the number of the unknown parameters, the combined ones and the pre-reduced ones, respectively. In complete solutions, there is of course only one u.

For the inter-technique combination (VLBI, SLR, GNSS, and DORIS) of the DGFI@TUM combination center, the VLBI input has to contain NUMBER OF OBSERVATIONS including the constraint equations for clocks and atmospheres, NUMBER OF UNKNOWNS, and WEIGHTED SQUARE SUM OF O-C (M. Seitz, priv. comm.)

14.3. Combination steps

In simple terms, the combination of pre-reduced normal equation systems consists of adding the individual entries in a weighted fashion and a subsequent inversion of the resulting system. Before combination, it has to be made sure that the normal equation systems are based on the same a priori values for the parameters to be combined. This is done with an 'a priori value transformation' (Vennebusch et al., 2007). The second transformation is applied to guarantee that the reference epochs for time variable parameters, such as EOP, are identical. This is done by 'epoch transformations'. For parameter transformations see also Brockmann (1997).

The addition of normal equation systems has to make sure that the elements, respectively the lines and columns, of the normal equations are added in correct order. Sometimes, this is also called stacking of normal equation systems. The mathematical formulation is

$$\Big(\sum_{i=1}^{p} \alpha_i \, \mathbf{A}_i^T \, \Sigma_{yy}^{-1} \, \mathbf{A}_i\Big)\mathbf{x}_1 = \sum_{i=1}^{p} \alpha_i \, \mathbf{A}_i^T \, \Sigma_{yy}^{-1} \, \mathbf{y}_i \tag{14.11}$$

which is

$$\Big(\sum_{i=1}^{p} \alpha_i \mathbf{N}_i\Big)\mathbf{x}_1 = \sum_{i=1}^{p} \alpha_i \mathbf{b}_i \tag{14.12}$$

where α_i are individual weight factors which can be derived in various forms, one of them being variance component estimation (Vennebusch et al., 2007). Care has to be taken that the correct lines and columns of the $\mathbf{N}_i$ are added for every parameter x. Summations are also carried out for

$$(\boldsymbol{\ell}^T \mathbf{P} \boldsymbol{\ell})_{sum} = (\mathbf{y}^T \, \boldsymbol{\Sigma}_{yy}^{-1} \, \mathbf{y})_{sum} = \sum_{i=1}^{p} (\mathbf{y}^T \, \boldsymbol{\Sigma}_{yy}^{-1} \, \mathbf{y})_{red_i} \qquad n_{sum} = \sum_{i=1}^{p} n_i \qquad u_{sum} = \sum_{i=1}^{p} u_i \tag{14.13}$$

to write these into the SINEX file of the combined solution for further combination with other techniques.

It may happen that the $(\mathbf{y}^T \, \boldsymbol{\Sigma}_{yy}^{-1} \, \mathbf{y})_{red}$ is not reported correctly which is currently the case for the Calc/Solve input to the IVS combination. If the square sum of residuals $\mathbf{v}^T \, \boldsymbol{\Sigma}_{yy}^{-1} \, \mathbf{v}$ and the respective parameter $\mathbf{x}$ and $\mathbf{b}$ vectors are available, a replacement can then be calculated with

$$\boldsymbol{\ell}^T \mathbf{P} \boldsymbol{\ell} = \mathbf{y}^T \, \boldsymbol{\Sigma}_{yy}^{-1} \, \mathbf{y} = \mathbf{v}^T \, \boldsymbol{\Sigma}_{yy}^{-1} \, \mathbf{v} + \mathbf{x}^T \, \mathbf{b} \tag{14.14}$$

Let us now consider the situation when we want to compute a solution from the VLBI pre-reduced normal equation systems for a whole suit of sessions up to a full global solution. For a single session alone, for which the logic of step-wise solutions is applied, the sequence of computational steps continues with datum definition and Eq. 14.15 directly. For the multi-session solution, we have to do the stacking of Eq. 14.12 in the same way as for single session combinations matching the order of the elements according to the vector of parameters.

Since the accumulation of multiple sessions necessarily extend the time base of the solution, velocity or higher order time derivatives have to be taken into account. This is done by extending the normal equation system by the respective elements for these parameters (Brockmann, 1997). The resulting normal equation system is still free of datum and only after datum definition as described in Sec. 13.8, the vector of the estimated parameters $\tilde{\mathbf{x}}_1$ can be computed with

$$\tilde{\mathbf{x}}_1 = \bar{\mathbf{N}}^{-1}\,\bar{\mathbf{b}}. \tag{14.15}$$

For completeness, it should be stated that the pre-reduced parameters of a session $\mathbf{x}_2$ can be computed in a back solution after the parameters $\mathbf{x}_1$ have been determined. The formula for this is

$$\tilde{\mathbf{x}}_2 = \mathbf{N}_{22}^{-1}\,(\mathbf{b}_2 - \mathbf{N_{21}}\,\tilde{\mathbf{x}}_1). \tag{14.16}$$

The respective cofactor matrix $\mathbf{Q}_{\tilde{x}_2\tilde{x}_2}$ is

$$\mathbf{Q}_{\tilde{x}_2\tilde{x}_2} = \mathbf{N}_{22}^{-1} + \mathbf{N}_{22}^{-1}\,\mathbf{N}_{21}\,(\mathbf{N}_{11} - \mathbf{N}_{12}\,\mathbf{N}_{22}^{-1}\,\mathbf{N}_{12})^{-1}\,\mathbf{N}_{12}\,\mathbf{N}_{22}^{-1} \tag{14.17}$$

or directly using the cofactor matrix of the combined parameters $\mathbf{Q}_{11}$

$$\mathbf{Q}_{\tilde{x}_2\tilde{x}_2} = \mathbf{N}_{22}^{-1} + \mathbf{N}_{22}^{-1}\,\mathbf{N}_{21}\,\mathbf{Q}_{11}\,\mathbf{N}_{12}\,\mathbf{N}_{22}^{-1} \tag{14.18}$$

The statistics also require to compute the variance of unit weight of each individual observing session to allow for the determination of the formal errors of the pre-reduced local parameters. For this purpose, we re-insert the estimated parameters $\tilde{\mathbf{x}}_1$ and $\tilde{\mathbf{x}}_2$ in the structure of Eq. 13.95 to compute the residuals $\mathbf{v}$

$$\mathbf{v} = \mathbf{A}\begin{pmatrix} \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_2 \end{pmatrix} - \mathbf{y} \tag{14.19}$$

where the product of $\mathbf{A}\cdot\tilde{\mathbf{x}}$ yields the "computed" observations $\tilde{\mathbf{y}}$. Then, the Sum of squares of residuals $\tilde{\Omega}$ is

$$\tilde{\Omega} = \mathbf{v}^T\,\mathbf{\Sigma}_{yy}^{-1}\,\mathbf{v} \tag{14.20}$$

and the variance of unit weight of the session is

$$\tilde{\sigma}_0^2 = \frac{\tilde{\Omega}}{(n+n_c)-(u_1+u_2)} \tag{14.21}$$

where n is the number of VLBI observations, n_c the number of constraints and u_1 is the number of combined parameters and u_2 that of the pre-reduced parameters. Finally, the covariance matrix $\mathbf{\Sigma}_{\tilde{x}_2\tilde{x}_2}$ is computed with

$$\mathbf{\Sigma}_{\tilde{x}_2\tilde{x}_2} = \tilde{\sigma}_0^2\,\mathbf{Q}_{\tilde{x}_2\tilde{x}_2}. \tag{14.22}$$

Although the use of Eq. 14.19 seems to be straight forward, it implies an adverse consequence. It requires that the **A** matrix is either available from the first reduction step or it needs to be recomputed which is computationally expensive.

15. International VLBI Service for Geodesy and Astrometry

15.1. History

Until the end of the last century, most observational and scientific activities in geodetic and astrometric VLBI were based on individual agreements of cooperation which required tedious multilateral communications. In addition, technical incompatibilities between recording and correlator systems were a major hindrance for more rapid progress (Schlüter and Behrend, 2007). At the same time, geodesy experienced major changes with respect to its organizational structure and its role within the global geoscientific community witnessing a shift to more global structures. On the initiative of the International Association of Geodesy (IAG), a number of technique-specific services were founded. Among them was the International VLBI Service for Geodesy and Astrometry (IVS) for coordinating the global VLBI resources and observing plans. The IVS was formally established on March 1, 1999, following the first meeting of the IVS Directing Board (DB) on February 11, 1999. More than a year prior to this date, a steering committee started work on the IVS Terms of Reference (ToR), a call for participation and the election process for the DB. Details and names of the persons involved can be found on the IVS Web page under https://ivscc.gsfc.nasa.gov/about/org/documents/index.html. Soon after its establishment, the IVS was recognized as a service of the International Astronomical Union (IAU) and of the Federation of Astronomical and Geophysical Data Analysis Services (FAGS); the latter was dissolved in 2009 and replaced by the World Data System (WDS) in 2011 (Behrend, 2013).

15.2. Duties

The IVS is an international collaboration of organizations, which operate or support VLBI components, for streamlining and improving a number of organisational and scientific issues. The IVS supports geodetic and astrometric work on reference systems and Earth science research in general, and provides the basis to all operational VLBI activities. Participation is organized on a voluntary proposal basis. According to the *IVS Terms of Reference* [99] its mission objectives are:

1. to foster and carry out VLBI programs;
2. to promote research and development activities in all aspects of the geodetic and astrometric VLBI technique;
3. to advance the education and training of VLBI participants through workshops, reports, and other means;

[99] http://ivscc.gsfc.nasa.gov/about/org/documents/ivsTOR.html

4. to support the integration of new components into IVS;

5. to interact with the community of users of VLBI products.

All of this is accomplished through close coordination of the participating organizations to provide high-quality VLBI data and products. To meet the objectives, IVS coordinates VLBI observing programs, sets performance standards for VLBI stations, establishes conventions for VLBI data formats and data products, issues recommendations for VLBI data analysis software, sets standards for VLBI analysis documentation, and institutes appropriate VLBI product delivery methods to ensure suitable product quality and timeliness. IVS also coordinates its activities with the astronomical community because of the dual use of many VLBI facilities and technologies for both radio astronomy and geodesy/astrometry.

IVS is also in charge of the integration of VLBI into a global Earth observing system and thus represents VLBI in the Global Geodetic Observing System (GGOS) of the IAG. IVS also interacts closely with the International Earth Rotation and Reference Systems Service (IERS), which is tasked by IAU and IUGG (International Union of Geodesy and Geophysics) with maintaining the reference frames ICRF and ITRF as described in Secs. 13.1.1 and 13.1.2.

15.3. Organizational structure

Strategic decisions and long-term plans of the IVS are governed by the IVS Directing Board (DB) composed of elected and ex-officio members. It meets twice a year in person or more often online if needed. The day-to-day activities are coordinated by the central bureau of the service, called IVS Coordinating Center (CC), which has been hosted by NASA's Goddard Space Flight Center in Greenbelt, MD, USA, since 1999.

Within the DB, three IVS Coordinators take responsibility for specific fields of the IVS's mandate. The Network Coordinator, beyond other tasks referring to the IVS network, "monitors adherence to standards in the network operation, participates in the quality control of the data acquisition performance of the network stations, tracks data quality and data flow problems, ... and coordinates software development for station control and monitoring."[99]

The Technology Coordinator is tasked, among others, with stimulating advancements of the VLBI technique, coordinating development of new technology among the various IVS Technology Development Centers, and encouraging technical compatibility within the IVS and with the astronomical community.

The Analysis Coordinator "fosters comparisons of results from different VLBI analysis software packages and different analysis strategies, ..., ensures that IVS analysis and combination products are archived and are available to the scientific community, and supervises the formation of the official IVS products specified by the IVS Directing Board."[99]

For the purpose of avoiding unnecessary ambiguities in data level descriptions, the IVS DB has resolved a "Nomenclature for data types and analysis steps" (IVS Resolution IVS-Res-2019-02): "... This nomenclature consists of the following dividers: *Level-0 Data* is the raw digitized noise gathered at the radio telescopes. Consequently, the correlation process falls under *Level-0 Data Analysis*. The output of this processing step are the fringe visibilities, subsumed as *Level-1 Data*. The analysis steps needed to produce the observables phase and group delays as well as their time derivatives, such as polarization combination and fringe fitting of the visibilities including any other necessary analysis steps at this stage be called *Level 1 Data Analysis*. The output is the *Level-2 Data*. The analysis steps working with phase and group delays and their rates and resulting in geodetic parameters be called *Level-2 Data Analysis*. This also includes work on source imaging and source structure effects. The results of this are geodetic and astrometric parameters forming *Level-3 Data*. The final combination work of several *Level-3 Data*, e.g., from different analysis centers, is *Level-3 Data Analysis*."

For specific purposes, the IVS Directing Board also proposes and installs IVS Working Groups[27] (WG). These are established on the basis of a specific charter and have a limited lifetime ending with the submission of a final report to be accepted by the DB. So far, seven WGs had been installed and finalized.

WG 1 GPS Phase Center Mapping

WG 2: IVS Product Specification and Observing Programs

WG 3: VLBI2010

IERS/IVS WG: Second Realization of the ICRF

WG 4: VLBI Data Structures

WG 5: Space Science Applications

WG 6: Education and Training

WG 7: Satellite Observations with VLBI

WG 8: Galactic Aberration

WG3 laid the groundwork for the VLBI2010 investigations, now called the VGOS concept with its new technology (Sec. 6). WG3 as well as WG7 have been converted to VGOS Technical Committee and Committee on Education and Training, respectively.

IVS Committees[28] are permanent organisational elements of the IVS handling specific tasks in close cooperation with the IVS Directing Board, the Coordinators and the Coordinating Center. The most important organisational entity within the IVS and below the IVS DB is the Observing Program Committee (OPC). Every year, with support of the CC, it prepares the IVS Master Schedule which determines the allocation of IVS resources into

[27] https://ivscc.gsfc.nasa.gov/about/wg/index.html
[28] https://ivscc.gsfc.nasa.gov/about/com/index.html

IVS observing programmes and individual observing sessions. The OPC meets regularly every month since imminent developments require timely responses. This also applies to proposals to the IVS for special types of observing sessions. The fourth IVS Committee is the Celestial Reference Frame Committee with emphasis on the continuous IVS astrometry efforts.

The IVS currently has more than 80 permanent components supported by more than 40 institutions in about 20 countries. The permanent components and their functions within the IVS are:

- *34 Network Stations:* acquiring high performance VLBI data;
- *3 Operation Centers:* coordinating the activities of specific networks of IVS radio telescopes;
- *7 Correlators:* processing the acquired data, providing feedback to the stations and providing processed data to analysts;
- *3 Data Centers:* distributing products to users, providing storage and archiving functions; a mirroring function guarantees failsafe operations of the IVS;
- *30 Analysis Centers:* analyzing the data and producing the results and products;
- *6 Technology Development Centers:* developing new VLBI technology;
- *1 Coordinating Center:* coordinating daily and long-term activities.

The observational network scheduled in the IVS observing plan comprises ∼30 network stations plus ∼15 stations from astronomical networks such as the VLBA (Very Large Baseline Array), the DSN (Deep Space Network), or the EVN (European VLBI Network), which cooperate with the IVS. A master observing plan is prepared for each calendar year based on the station time available for IVS observations. Some stations carry a high load of observations and are included in most of the observing sessions, while other stations can only contribute to dedicated campaigns (Schlüter and Behrend, 2007). In 2021, there are three-and-a-half 24-hour sessions and seven daily 1-hour intensive sessions carried out per week. In other words, there are VLBI observations available for roughly 50% of the year.

The central hub(s) of the IVS data flow are the IVS Data Centers (Fig. 15.1). Here, the observing schedules are stored for the observatories to retrieve them prior to the respective date of the observing session. After the observation session has been completed, the telescope data is transferred to a correlator center assigned for the correlation task in the master schedule. For setting up the correlator control files, the observing schedule is retrieved from the data centers as well. The results of the correlation and fringe fitting process (*Level 1* and *Level 2 Data*) are then stored at the data centers for the analysis centers to retrieve them for their processing. The results again go back to the data centers in agreed-upon formats. One of these formats is the so-called SINEX format (Solution Independent

Exchange Format)[29] This is used by the IVS Combination Center to produce a combined solution for each observing session which is more robust against individual blunders and, in most cases, produces better results than the individual ones.

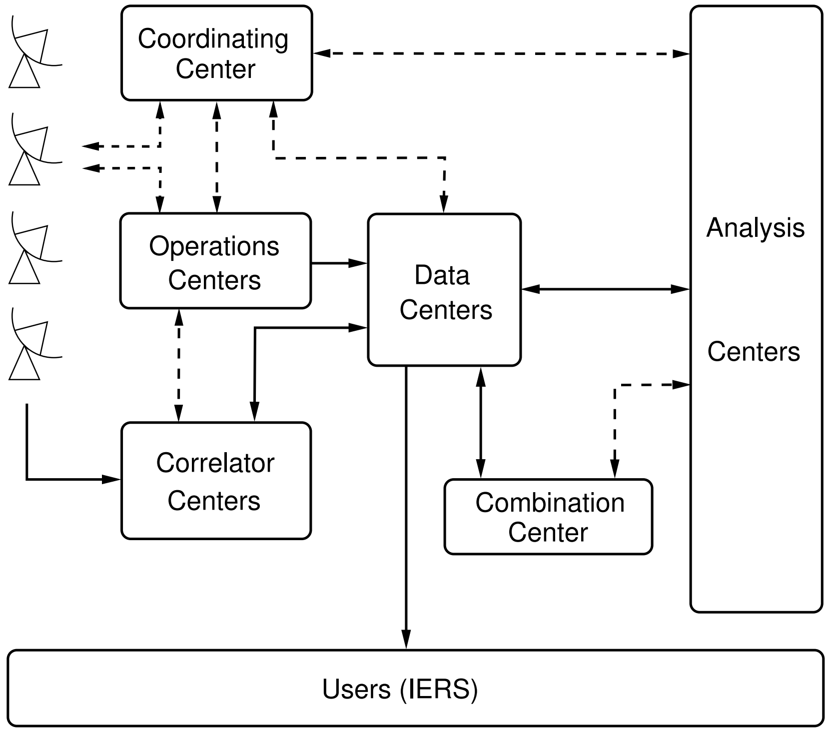


Figure 15.1: IVS Components with major data flow links (solid lines) and information exchange links (dashed lines)

.

15.4. Current Applications and Products

Under the umbrella of the IVS, geodetic and astrometric VLBI is being applied to furnish a number of products. These are provided on a regular basis and stem from a continuous monitoring program. VLBI data products currently available are the full set of Earth orientation parameters, the terrestrial reference frame (TRF), and the celestial reference frame (CRF), (see also Sec. 16). The IVS products can be defined in terms of their accuracy, reliability, frequency of observing sessions, temporal resolution of the estimated parameters, time delay from observing to final product, and frequency of solutions. All VLBI data and results in appropriate formats are archived in IVS Data Centers and are publicly available for research in related areas of geodesy, geophysics, and astrometry. The IVS data set extends from 1979 to date with new results being made available with some latency needed for the processing (Table 16.1).

[29]https://www.iers.org/IERS/EN/Organization/AnalysisCoordinator/SinexFormat/sinex.html

More details of the IVS, in particular of the observing programme can be found in Schlüter and Behrend (2007), Schuh and Behrend (2012), and Nothnagel et al. (2017) as well as online at http://ivscc.gsfc.nasa.gov.

16. Results

Session-wise results are of particular interest in the case of Earth orientation parameters. Unique for VLBI are the results of the Earth's phase of rotation UT1, which cannot be determined with such a quality and time resolution with any other technique in particular not with any satellite technique. The reason is that UT1 is directly correlated with the ascending nodes of any satellites' orbits. The phase of rotation is reported as time difference of UT1 – UTC. Formal errors are at the level of 2-3 μs for contemporary network sessions while daily single baseline sessions of one hour duration only (so-called *Intensives*) provide UT1–UTC with a standard deviation of about 8 -15 μs.

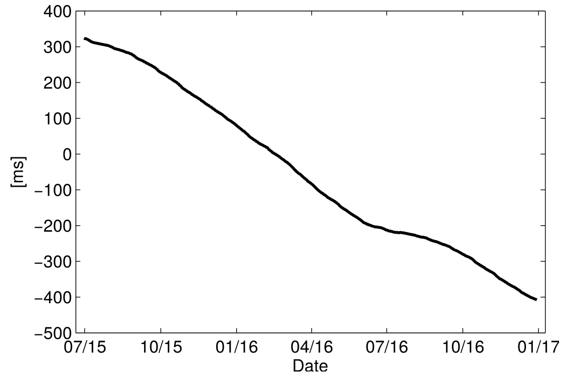


Figure 16.1: Results of UT1-UTC between two epochs where leap seconds were introduced.

Unique are also the estimated adjustments to the precession/nutation model. They are accurate to about 80 μas and represent not so much inaccuracies in the precession/nutation model but variations caused by free core nutation (Dehant and Mathews, 2015) which are at the level of about 1 mas.

Daily polar motion estimates provide a reliable time series of the long term evolution of polar motion (Fig. 16.2). Although the formal errors are slightly worse than those of the IGS polar motion time series, they are needed for cross-validation.

Table 16.1: Synopsis of the main products of the IVS (based on Schlüter and Behrend (2007)) and their specifications at the time of writing. Estimates for UT1—UTC are determined on the basis of the 24-hour sessions as part of the full set of five EOP parameters and on the basis of the 1-hour Intensive sessions for UT1 prediction purposes.

	Polar motion (x_p, y_p)	UT1—UTC 24-hr session	UT1—UTC Intensive	Celestial pole (dX, dY)
Accuracy	50–80 μas	3–5 μs	15–20 μs	50 μas
Product delivery	8–10 days	8–10 days	1 day	8–10 days
Resolution	1 day	1 day	1 day	1 day
Frequency of solution	~3 days/week	~3 days/week	daily	~3 days/week

Telescope coordinates and their derivatives with respect to time, the so-called veloci-

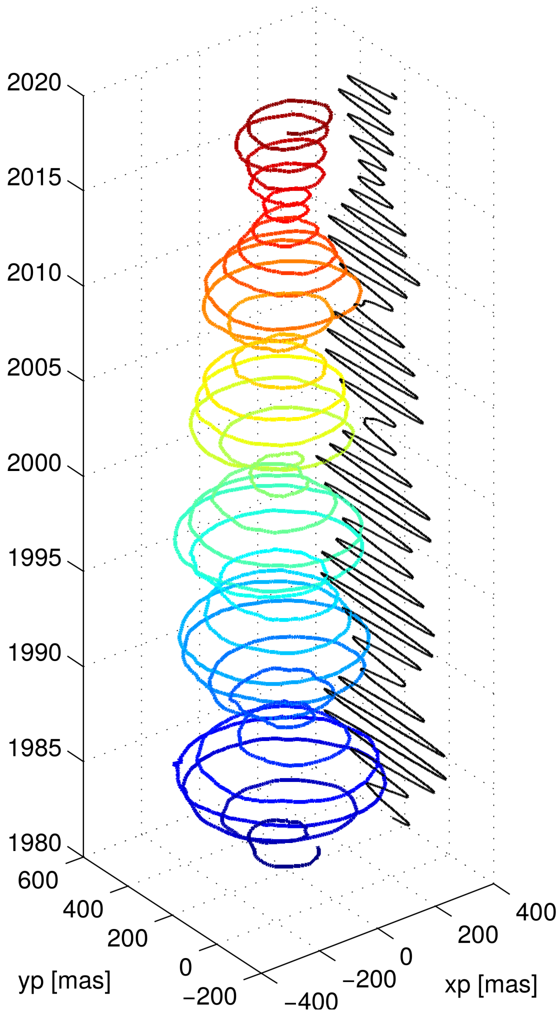


Figure 16.2: Long term polar motion.

Table 16.2: TRF and CRF products of the IVS (based on Schlüter and Behrend 2007).

	TRF (x, y, z)	CRF (α, δ)
Accuracy	5 mm	40–250 μas
Product delivery	—	3 months
Resolution	—	—
Frequency of solution	—	1 year

ties, are always estimated in geocentric Cartesian coordinates in the International Terrestrial Reference System (Sec. 13.1.1). Since VLBI has six degrees of freedom (3 translations and 3 rotations), the datum always needs to be defined for any VLBI solution. Most commonly existing coordinates and velocities tabulated in the International Terrestrial Reference Frame in its latest versions are used for this purpose. To avoid dependencies on individual telescopes and their observing history and possible peculiarities, so called no-net-rotation (NNR) and no-net-translation (NNT) conditions or constraints with respect to this or any other reference frame are jointly applied to define the datum.

For many purposes it is convenient to express the VLBI results in local topocentric systems. This allows to investigate, for example, the quality of the results in the local vertical direction which is always the component determined with less quality than the horizontal components. For details on the x,y,z to Up, East, North components please see Sec. A.2

Baseline lenghts

From the time series of telescope coordinates, baseline lengths can be inferred. These are independent of any datum and can conveniently be used for stability investigations. One of the most prominent baseline is that between the 18 m telescope of MIT Haystack Observatory near Westford (MA) at the East coast of the U.S. and the 20 m telescope of the Bundesamt für Kartographie und Geodäsie and the Technische Universität München near Wettzell, Germany (Fig. 16.3). It should be noted that the input to the time series is not constrained in any way and still the approximation by a linear regression is extraordinarily precise. Fig. 16.3 has been produced by a Web tool of the IVS under https://www.ccivs.bkg.bund.de/index.php?uri=quarterly/baseline where plots of all VLBI time series of baseline lengths are available interactively.

Results originating from VLBI global solutions guarantee consistency between parameters estimated in the same solution setup. Standard parameters here are radio source positions, which are being used in the construction of the ICRF (Charlot et al., 2020). Telescope coordinates and velocities regularly enter the computations of the specific realizations of the International Terrestrial Reference System (ITRS), the ITRF (Altamimi et al., 2016; Bachmann et al., 2016). The standard deviations of the coordinates of the most reliable and often used telescopes are at the level of about 1 – 3 mm with velocity components

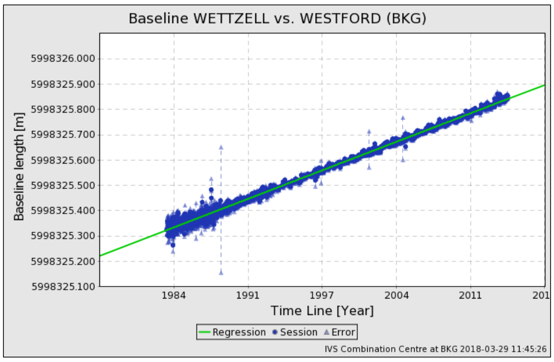


Figure 16.3: Time series of baseline lengths between the Wettzell and Westford radio telescopes with linear regression line.

having standard deviations of 0.1 mm/y.

Appendices

A. Rotations

A.1. Fundamental rotations

Transformations of points, vectors or tripods are almost always accompanied by rotations. For the correct application of the mathematical formulations, it is important to distinguish between the purposes of the rotations. In many technical cases, rotations are applied to a body, most easily represented by a vector, which will have a different orientation in the rotated position. In such a case, the coordinate system remains the same and the rotations are called *Category 1* rotations.

In the second case, called *Category 2*, the object virtually remains fixed but the axes of the system(s) are rotated. This case applies to transformations of coordinates from one system to another. As we will see, there are fundamental differences between the two categories.

Any 3D-rotations can always be disassembled into three fundamental rotations around the respective axes. Important is the definition of the signs of the rotation angles because it determines how the rotation matrices below are set up. In mathematics, the rotations are considered positive, if the rotation is clockwise as seen from the origin (Fig. A.1). As seen along the axis towards the origin, these rotations are, of course, negative.

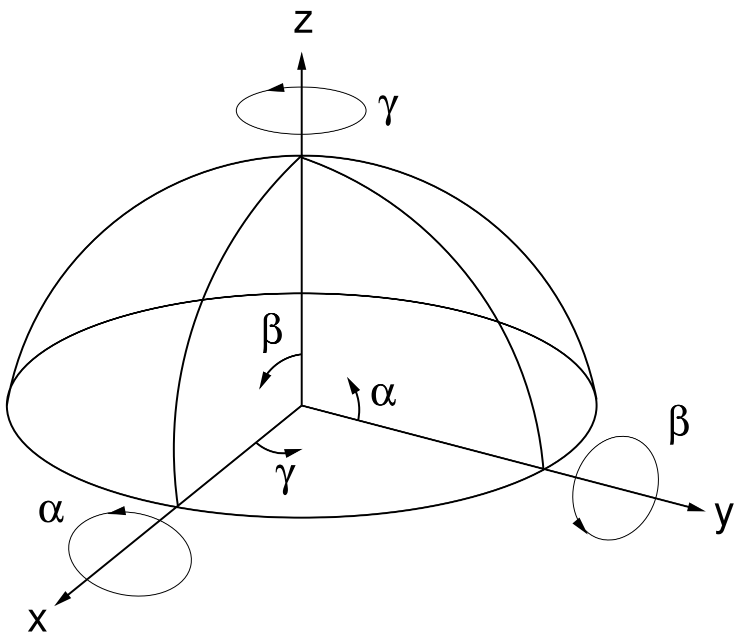


Figure A.1: Directions of fundamental rotations in a mathematical sense.

The rotation of any vector $\mathscr{X}$ in the x, y, z system around an angle is performed through the application of a rotation matrix $\mathbf{R'}_*$ into a vector in the rotated system $\mathscr{X}'$ by

$$\mathscr{X}' = \mathbf{R'}_* \cdot \mathscr{X} \tag{A.1}$$

where "*" is an index with $* \in \{1,2,3\}$ indicating around which axis the rotation is performed. The rotation matrices $\mathbf{R'}_*$ of any elementary rotations α around the 1-axis have the form

$$\mathbf{R'}_1(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \tag{A.2}$$

while for rotations around the 2- and 3-axis with β and γ, respectively, look very similar:

$$\mathbf{R'}_2(\beta) = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \tag{A.3}$$

$$\mathbf{R'}_3(\gamma) = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{A.4}$$

The prime attached to the rotation matrix indicates that this is according to the mathematical convention. To Carry out a sequence of rotations around all three axes, the second and third rotations are appended to Eq. A.1 "from the left" because the axes of the object move with the rotations.

$$\mathscr{X}' = \mathbf{R'}_3(\gamma) \cdot \mathbf{R'}_2(\beta) \cdot \mathbf{R'}_1(\alpha) \cdot \mathscr{X} = \mathbf{R'}(\alpha, \beta, \gamma) \cdot \mathscr{X} \tag{A.5}$$

The rotation matrix $\mathbf{R'}$ can also be composed a priori by the three rotation matrices. The sequence of rotations in Eq. A.1 is only an example and may be different depending on the application.

For *Category 2* rotations, the concept is slightly different. Here, we rotate the axes. However, any point, e.g., represented by a position vector, remains at its original location. This is the classical case of a coordinate transformation from $\mathbf{x}$ in the original system into $\mathbf{x}'$ in the rotated system. The rotations are now performed by the inverse or transposed rotation matrices ($\mathbf{R}^{-1} = \mathbf{R}^T$)

$$\mathbf{x}' = \mathbf{R'}^T_3(\gamma) \cdot \mathbf{R'}^T_2(\beta) \cdot \mathbf{R'}^T_1(\alpha) \cdot \mathbf{x} = \mathbf{R'}^T(\alpha, \beta, \gamma) \cdot \mathbf{x} \tag{A.6}$$

The three individual rotation matrices, which keep the initial sign convention of Eq. A.2, A.3, and A.4, are as follows

$$\mathbf{R}'^{T}_{1}(\alpha) = \mathbf{R}_1(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix} \tag{A.7}$$

$$\mathbf{R}'^{T}_{2}(\beta) = \mathbf{R}_2(\beta) = \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix} \tag{A.8}$$

$$\mathbf{R}'^{T}_{3}(\gamma) = \mathbf{R}_3(\gamma) = \begin{pmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{A.9}$$

The three rotation matrices in Eqs. A.7, A.8, and A.9 are now void of the prime and the transposed symbol because these are the rotation matrices found in most astronomical text books (e.g. Dehant and Mathews 2015) which mostly deal with coordinate transformations. Therefore, they are also used in the same way in this tutorial.

A.2. Rotations from geocentric into topocentric system

For many purposes it is convenient to express the VLBI results in local topocentric systems. This allows to investigate, for example, the quality of the results in the local vertical direction. These topocentric systems normally are right hand systems defined with up, east and north (U,E,N) components with the up component being the local vertical and the north component pointing towards the north coordinate direction in the tangential plane onto the Earth's surface. The East axis completes the right hand system perpendicular to the up-north plane. Due to the curvature of the Earth, the U,E,N representation is limited to the close vicinity of the telescope. For this reason, a reference point needs to be defined which most commonly is the point located at the reference coordinates (x_0, y_0, z_0) of the telescope. Displacements or velocity vectors (Δx, Δy, Δz or v_x, v_y, v_z) in the geocentric system can then be converted to the local topocentric one by applying the respective rotations in longitude λ and latitude ϕ. This is a *Cathegory 2* rotation since the tripod of deltas or velocities is a 3D vector in the xyz system which is kept fixed in its orientation. For the transformations, the coordinate axes are rotated by λ and ϕ according to Eqs. A.9 and A.8.

The peculiarities of these rotations can best be described in two individual steps. We start with the tripod of displacements (Δx, Δy, Δz) in an arbitrary telescope position (Fig. A.2). For an intuitive visualisation of the derivation of the rotations, it is convenient to re-locate this tripod to the intersection of the equator and the zero meridian. From here, we first rotate the tripod around the 3-axis from the zero longitude to the longitude of the telescope λ. By convention λ increases to the East which is counterclock-wise and thus positive in

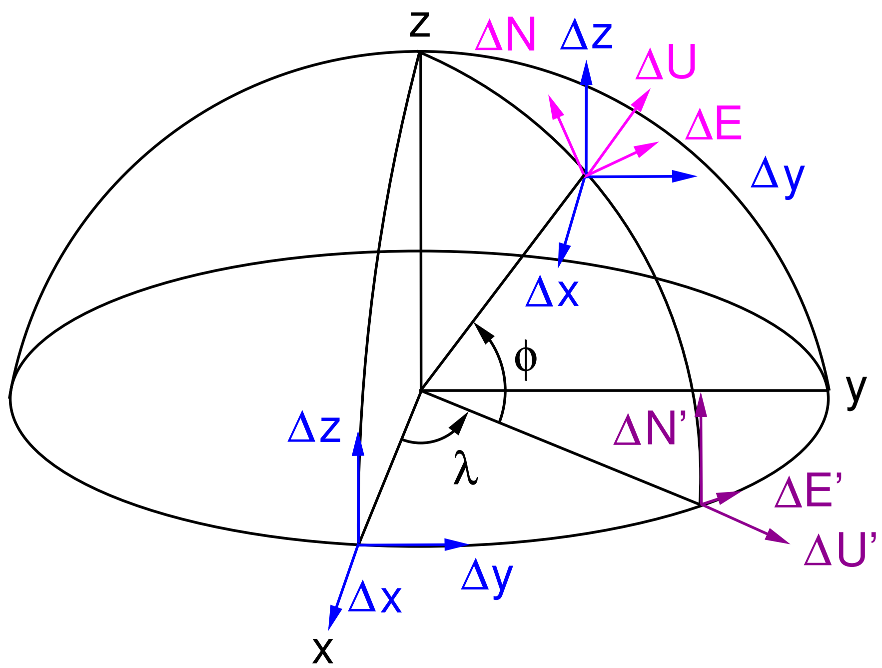


Figure A.2: Rotations of telescope displacements in geocentric system (Δx, Δy, Δz) into local topocentric system ($\Delta U(p)$, $\Delta N(orth)$, $\Delta E(ast)$).

the convention of the rotation matrix (Eq. A.9). We can thus write

$$\begin{pmatrix} \cos\lambda & \sin\lambda & 0 \\ -\sin\lambda & \cos\lambda & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \begin{pmatrix} \Delta x\cdot\cos\lambda + \Delta y\cdot\sin\lambda \\ -\Delta x\cdot\sin\lambda + \Delta y\cdot\cos\lambda \\ \Delta z \end{pmatrix} = \begin{pmatrix} \Delta U' \\ \Delta E' \\ \Delta N' \end{pmatrix}. \quad \text{(A.10)}$$

The second rotation around axis #2 with the latitude ϕ is clockwise and thus negative in the convention of Eq. A.8 requiring a swap of the sign. This can either be done by multiplying ϕ by -1 or by swapping the signs of the two sin elements in Eq. A.8 (identical to a transposition). For the latter, we then find

$$\begin{pmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{pmatrix} \begin{pmatrix} \Delta U' \\ \Delta E' \\ \Delta N' \end{pmatrix} = \begin{pmatrix} (\Delta x\cdot\cos\lambda + \Delta y\cdot\sin\lambda)\cdot\cos\phi + \Delta z\cdot\sin\phi \\ -\Delta x\cdot\sin\lambda + \Delta y\cdot\cos\lambda \\ -(\Delta x\cdot\cos\lambda + \Delta y\cdot\sin\lambda)\cdot\sin\phi + \Delta z\cdot\cos\phi \end{pmatrix} = \begin{pmatrix} \Delta U \\ \Delta E \\ \Delta N \end{pmatrix} \quad \text{(A.11)}$$

From Eq. A.11 we can now also extract the pure composite rotation matrix by isolating the Δx, Δy, and Δz because we can use this for the computation of the formal errors of the local components through variance propagation:

$$\mathbf{R}(\lambda,\phi) = \begin{pmatrix} \cos\lambda\cdot\cos\phi & \sin\lambda\cdot\cos\phi & \sin\phi \\ -\sin\lambda & \cos\lambda & 0 \\ -\cos\lambda\cdot\sin\phi & -\sin\lambda\cdot\sin\phi & \cos\phi \end{pmatrix} \quad \text{(A.12)}$$

However, before we come to this, on a sideline we briefly look at the direct multiplication of the rotation matrices. Although the $\mathbf{R}_3$ rotation around λ is the first one to be applied and the rotation $\mathbf{R}_2$ around ϕ comes second, the sequence is reversed, i.e., applied "from the left", because the axes move with the rotations. Longitude λ and latitude ϕ are applied as defined which is positive to the East and positive to the North, respectively (note the "transposed" of the $\mathbf{R_2}$ matrix):

$$\mathbf{\Delta}x_{UEN} = \mathbf{R}_2^T(\phi) \cdot \mathbf{R}_3(\lambda) \cdot \mathbf{\Delta}x_{xyz} \tag{A.13}$$

$$\mathbf{R}(\lambda, \phi) = \begin{pmatrix} \cos\lambda \cdot \cos\phi & \sin\lambda \cdot \cos\phi & \sin\phi \\ -\sin\lambda & \cos\lambda & 0 \\ -\cos\lambda \cdot \sin\phi & -\sin\lambda \cdot \sin\phi & \cos\phi \end{pmatrix} \tag{A.14}$$

This is of course identical to the step-by-step derivation of Eq. A.12.

A.3. Error propagation from geocentric to topocentric rotations

As we have seen, the topocentric components $\Delta(U)p$, $\Delta(E)ast$, and $\Delta(N)orth$, are not readily available in least squares adjustments of VLBI observing sessions since the model is normally represented in geocentric x,y,z coordinates. Hence, the variance/covariance matrix is also expressed geocentrically. To investigate the topocentric components and especially the vertical/up component and its correlations, the law of error propagation has to be applied to the covariance matrix $\mathbf{Q}$ in the form (Koch, 1999):

$$\mathbf{Q}^{UEN} = \mathbf{B} \cdot \mathbf{Q}^{xyz} \cdot \mathbf{B}^T \tag{A.15}$$

$\mathbf{B}$ is the matrix of partial derivatives with respect to the individual parameters or the so-called design matrix (here we only display $\mathbf{B}$ for a single site, for all other sites this pattern repeats):

$$\mathbf{B} = \begin{vmatrix} \frac{\partial U}{\partial x} & \frac{\partial U}{\partial y} & \frac{\partial U}{\partial z} & \frac{\partial U}{\partial CL0} & \frac{\partial U}{\partial CL1} & \frac{\partial U}{\partial CL2} & \frac{\partial U}{\partial AT1} & \cdots & \frac{\partial U}{\partial ATn} & \frac{\partial U}{\partial NG} & \frac{\partial U}{\partial EG} \\ \frac{\partial E}{\partial x} & \frac{\partial E}{\partial y} & \frac{\partial E}{\partial z} & \frac{\partial E}{\partial CL0} & \frac{\partial E}{\partial CL1} & \frac{\partial E}{\partial CL2} & \frac{\partial E}{\partial AT1} & \cdots & \frac{\partial E}{\partial ATn} & \frac{\partial E}{\partial NG} & \frac{\partial E}{\partial EG} \\ \frac{\partial N}{\partial x} & \frac{\partial N}{\partial y} & \frac{\partial N}{\partial z} & \frac{\partial N}{\partial CL0} & \frac{\partial N}{\partial CL1} & \frac{\partial N}{\partial CL2} & \frac{\partial N}{\partial AT1} & \cdots & \frac{\partial N}{\partial ATn} & \frac{\partial N}{\partial NG} & \frac{\partial N}{\partial EG} \\ \frac{\partial CL0}{\partial x} & \frac{\partial CL0}{\partial y} & \frac{\partial CL0}{\partial z} & \frac{\partial CL0}{\partial CL0} & \frac{\partial CL0}{\partial CL1} & \frac{\partial CL0}{\partial CL2} & \frac{\partial CL0}{\partial AT1} & \cdots & \frac{\partial CL0}{\partial ATn} & \frac{\partial CL0}{\partial NG} & \frac{\partial CL0}{\partial EG} \\ \frac{\partial CL1}{\partial x} & \frac{\partial CL1}{\partial y} & \frac{\partial CL1}{\partial z} & \frac{\partial CL1}{\partial CL0} & \frac{\partial CL1}{\partial CL1} & \frac{\partial CL1}{\partial CL2} & \frac{\partial CL1}{\partial AT1} & \cdots & \frac{\partial CL1}{\partial ATn} & \frac{\partial CL1}{\partial NG} & \frac{\partial CL1}{\partial EG} \\ \frac{\partial CL2}{\partial x} & \frac{\partial CL2}{\partial y} & \frac{\partial CL2}{\partial z} & \frac{\partial CL2}{\partial CL0} & \frac{\partial CL2}{\partial CL1} & \frac{\partial CL2}{\partial CL2} & \frac{\partial CL2}{\partial AT1} & \cdots & \frac{\partial CL2}{\partial ATn} & \frac{\partial CL2}{\partial NG} & \frac{\partial CL2}{\partial EG} \\ \frac{\partial AT1}{\partial x} & \frac{\partial AT1}{\partial y} & \frac{\partial AT1}{\partial z} & \frac{\partial AT1}{\partial CL0} & \frac{\partial AT1}{\partial CL1} & \frac{\partial AT1}{\partial CL2} & \frac{\partial AT1}{\partial AT1} & \cdots & \frac{\partial AT1}{\partial ATn} & \frac{\partial AT1}{\partial NG} & \frac{\partial AT1}{\partial EG} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ \frac{\partial ATn}{\partial x} & \frac{\partial ATn}{\partial y} & \frac{\partial ATn}{\partial z} & \frac{\partial ATn}{\partial CL0} & \frac{\partial ATn}{\partial CL1} & \frac{\partial ATn}{\partial CL2} & \frac{\partial ATn}{\partial AT1} & \cdots & \frac{\partial ATn}{\partial ATn} & \frac{\partial ATn}{\partial NG} & \frac{\partial ATn}{\partial EG} \\ \frac{\partial NG}{\partial x} & \frac{\partial NG}{\partial y} & \frac{\partial NG}{\partial z} & \frac{\partial NG}{\partial CL0} & \frac{\partial NG}{\partial CL1} & \frac{\partial NG}{\partial CL2} & \frac{\partial NG}{\partial AT1} & \cdots & \frac{\partial NG}{\partial ATn} & \frac{\partial NG}{\partial NG} & \frac{\partial NG}{\partial EG} \\ \frac{\partial EG}{\partial x} & \frac{\partial EG}{\partial y} & \frac{\partial EG}{\partial z} & \frac{\partial EG}{\partial CL0} & \frac{\partial EG}{\partial CL1} & \frac{\partial EG}{\partial CL2} & \frac{\partial EG}{\partial AT1} & \cdots & \frac{\partial EG}{\partial ATn} & \frac{\partial EG}{\partial NG} & \frac{\partial EG}{\partial EG} \end{vmatrix}. \tag{A.16}$$

Since in different coordinate systems there are only interdependencies in the coordinate components but not between other parameters, the design matrix can be simplified to read (again for one site only):

$$\mathbf{B} = \begin{vmatrix} \frac{\partial U}{\partial x} & \frac{\partial U}{\partial y} & \frac{\partial U}{\partial z} & 0 & \cdots & 0 \\ \frac{\partial E}{\partial x} & \frac{\partial E}{\partial y} & \frac{\partial E}{\partial z} & 0 & \cdots & 0 \\ \frac{\partial N}{\partial x} & \frac{\partial N}{\partial y} & \frac{\partial N}{\partial z} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & & 0 \\ \vdots & \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & 0 & & 1 \end{vmatrix}. \tag{A.17}$$

with

$$\begin{aligned} \frac{\partial U}{\partial x} &= \cos\phi\cos\lambda, & \frac{\partial U}{\partial y} &= \cos\phi\sin\lambda, & \frac{\partial U}{\partial z} &= \sin\phi \\ \frac{\partial E}{\partial x} &= -\sin\lambda, & \frac{\partial E}{\partial y} &= \cos\lambda, & \frac{\partial E}{\partial z} &= 0 \\ \frac{\partial N}{\partial x} &= -\sin\phi\cos\lambda, & \frac{\partial N}{\partial y} &= -\sin\phi\sin\lambda, & \frac{\partial N}{\partial z} &= \cos\phi \end{aligned} \tag{A.18}$$

The correlation coefficients r_{ij} can then be computed from the rotated variance/covariance matrix, i.e., from the main diagonal elements (standard deviations $\sigma_U = \sqrt{Q_{UU}}$, $\sigma_E = \sqrt{Q_{EE}}$ and $\sigma_N = \sqrt{Q_{NN}}$) and the corresponding off-diagonal elements, i.e., the covariances Q_{ij}:

$$r_{ij} = \frac{Q_{ij}}{\sqrt{Q_{ii} \cdot Q_{jj}}}. \tag{A.19}$$

A different route to the same result for the rotated covariance matrix is the derivation with arrays. Here we start with the composed rotation matrix $\mathbf{R}(\lambda, \phi)$ to be applied to the $\mathbf{\Delta}x$

vector

$$\mathbf{\Delta}x_{UEN} = \mathbf{R}(\lambda, \phi) \cdot \mathbf{\Delta}x_{xyz} \tag{A.20}$$

If we write the derivatives in vector form, we find that

$$\frac{\partial \mathbf{\Delta}x_{UEN}}{\partial \mathbf{\Delta}x_{xyz}} = \frac{\partial(\mathbf{R}(\lambda, \phi) \cdot \mathbf{\Delta}x_{xyz})}{\partial \mathbf{\Delta}x_{xyz}} = \mathbf{R}(\lambda, \phi) \tag{A.21}$$

This then simplifies the error propagation of Eq. A.15 to

$$\mathbf{Q}^{UEN} = \mathbf{R}(\lambda, \phi) \cdot \mathbf{Q}^{xyz} \cdot \mathbf{R}^{T}(\lambda, \phi) \tag{A.22}$$

which is in perfect agreement with the individual partial derivatives in Eq. A.18.

B. Parallactic angle

For compensating for feed rotation effects in phase delay or total phase solutions, the parallactic angle ψ_{RS} is needed. It is the apparent orientation of the feed horn and is defined as the angle between the celestial meridian of the radio source and the meridian of the source in the topocentric (terrestrial telescope-centric) system (Thompson et al., 2007).

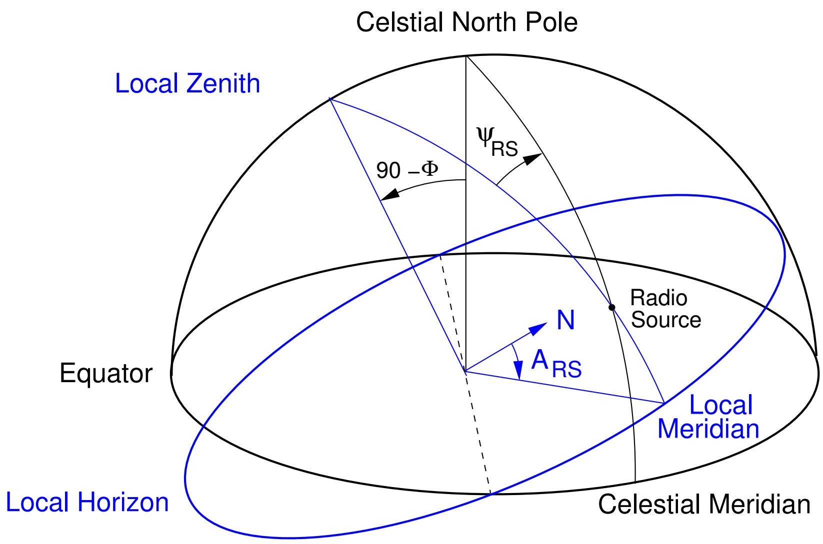


Figure B.1: Parallactic angle ψ_{RS} at the intersection of the celestial and topocentric meridians. Through the rotation of the local (blue) system eastward ψ_{RS} is negative in this scenario, turning positive after the radio source has culminated.

Using Eq. 9.1 for determining the local Hour Angle h_{loc}, Eq. 9.3 for the elevation of the radio source above the local horizon ε, and Eq. 9.2 for the azimuth of the radio source as seen from the telescope A_{RS}, allows for two separate ways to compute the parallactic angle ψ_{RS}. Through spherical trigonometry it can be derived that

$$\sin\psi_{RS} = \frac{\sin h_{loc}}{\cos\varepsilon} \cdot \cos\Phi. \tag{B.1}$$

or

$$\sin\psi_{RS} = \frac{-\sin A_{RS}}{\cos\delta} \cdot \cos\Phi. \tag{B.2}$$

Φ is the latitude of the telescope and δ the declination of the radio source.

As can be seen from Eq. B.1, ψ_{RS} is zero for observations along the local meridian being negative before the culmination and positive afterwards. This direction automatically serves as the reference for applying corrections to the observed phase for feed rotations.

C. Polarization of electromagnetic waves

C.1. General characteristics

Polarization is an important characteristic of electromagnetic waves at their origin but also while travelling through media and through VLBI equipment. The phenomenon was first studied and explained in the optical domain. However, the physical laws of optical wavelengths also apply in the microwave domain.

Electromagnetic waves are transversal waves. Polarization in general describes the plane of the electric field in which a transverse wave propagates. Under the condition that the electric field does not change its direction, the electric vector of a transmitted signal only oscillates within the polarization plane (red curve in Fig. C.1). The plane of the magnetic field, which is completely determined by the polarization of the electric field, is always perpendicular to the electric field, and both are normal to the propagation direction. The polarization of the magnetic field is determined by rotating the electric field 90° clock-wise looking in the direction of the propagation.

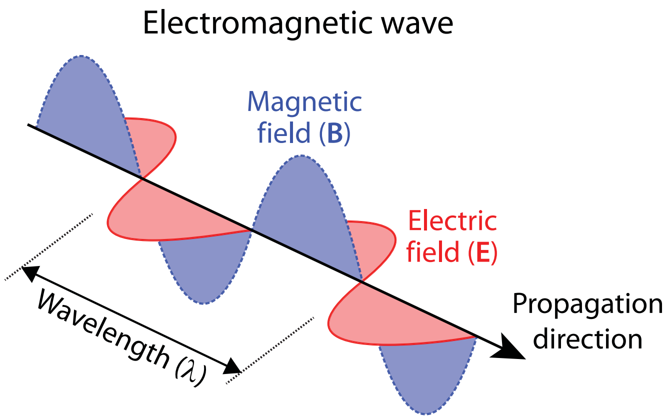


Figure C.1: Magnetic field and perpendicular electric field (Verhoeven, 2017)

.

If there is only one plane, we speak of linear polarization. In geodetic and astrometric VLBI, linear polarizations are important in the context of VGOS observations (see Sec. 6). Depending on the chosen reference frame, we can define for example vertical linear polarization (V), horizontal linear polarization (H), or a plane with any position angle. In the latter case, two perpendicular planes can be named X and Y. The naming depends on the chosen convention.

If we consider the electromagnetic radiation of natural radio sources, it consists of many concurrent wave trains. Here, the dominant electric field plane changes quickly and randomly. Then the wave is said to be unpolarized. The degree of polarization describes how much of the radiation follows a certain polarization plane. This is a statistical property of the aggregate wave trains.

Besides linear polarization, VLBI makes use of circular polarization. This type of polarization has been employed in the legacy style observations since the early days of VLBI in order to guarantee identical polarization at any telescope on Earth. Circular polarization originates from the fact that the electric vector of a transmitted signal traces the form of a circle (Thompson et al., 2007). Here, the magnitude of the vector is constant as is the angular velocity. Circular polarization can also be seen as the superposition of two perpendicular linearly polarized waves with the same frequency but with a constant phase difference of 90°. Depending on whether the phase shift is negative (-90°) or positive (90°), we speak of a right hand circular polarization (RCP) or a left hand circular polarization (LCP), respectively. If there is a phase offset between the two perpendicular sinusoidal waves, which is not (-90°) or (90°), the more general form is an elliptical polarization.

The polarization in VLBI first of all describes how a wave is received at the radio telescope and its receiving components such as the main and possibly sub-reflectors, the feed horn and the de-couplers. Of course the intrinsic polarization of emission from a radio source depends on its synchroton radiation and the magnetic field within the source. Synchroton radiation is caused by high-energy electrons in magnetic fields radiating as a result of their orbital motion (Thompson et al., 2007). Most compact extra-galactic radio sources, which we use in geodetic and astrometric VLBI, appear unpolarized with our synthesized resolution, which may be causing "beam smearing" of the polarization (If we had much higher resolutions, we might eventually detect polarization). The randomness allows to configure our VLBI equipment to receive only special polarization components such as linearly or circularly polarized radiation.

Astronomers, in a sub-field called polarimetry, are interested in radio sources with a non-zero state of polarization, which shows as non-zero cross-correlations between the orthogonal polarization components (X/Y, H/V or RCP/LCP)(I. Martí-Vidal, pers. comm.) since these help to explain physical phenomena and morphology of these sources. The difficulties caused by the fact, that telescopes separated by long baselines see the same source with a different and constantly changing orientation with respect to the telescope's axes, are overcome by employing dual circular polarization receivers (LCP and RCP). The combination of the two signals can be used to answer questions such as how much radiation is polarized vs. unpolarized, how much polarization is circular and in which direction, or whether there is distinct linear polarization and at what strength and orientation (Thompson et al., 2007). These questions can be answered through the use of the so-called Stokes parameters (Stokes (1852), see below).

In the context of propagation of the radiation from the source to the receiver, we should also mention Faraday rotation. Faraday detected that the plane of polarization rotates in a magnetic field. Here, the amount of rotation is proportional to the magnetic flux density, electron density (integrated along the path length) and the square of the wavelength. Hence, the detection of Faraday rotation from a source allows astronomers to probe the magneto-ionic medium (plasma with a magnetic field) in the emitting source providing a very important source of astrophysical information. However, Faraday rotation also occurs in the medium located within the line of sight to the Earth which includes the Earth's ionosphere.

C.2. Mathematical formulations

In mathematical terms, the electric field $\mathscr{E}$ of a monochromatic transverse wave within a plane wave front in the direction of a unit vector $\hat{\mathbf{a}}$ (Fig. C.2) can be described as (Burke et al., 2019)

$$\mathscr{E}(x,y) = \hat{\mathbf{a}}\,\mathscr{E}_0 \cos[2\pi(\nu t - kz) + \phi_0] \tag{C.1}$$

where the wave front propagates in the z direction with inverse wavelength $k = 1/\lambda$, frequency ν and phase offset ϕ_0. $\mathscr{E}_0$ is the amplitude of the electric field. In the complex form

$$\mathscr{E}(x,y) = \hat{\mathbf{a}}\,\mathscr{E}_0\, e^{i\,2\pi(\nu t - kz)} \tag{C.2}$$

the electric field $\mathscr{E}$, now contains the phase information. The polarization direction is defined by the behavior of $\hat{\mathbf{a}}$. If it has a constant orientation, as it is assumed in this derivation, the wave is linearly polarized (Burke et al., 2019).

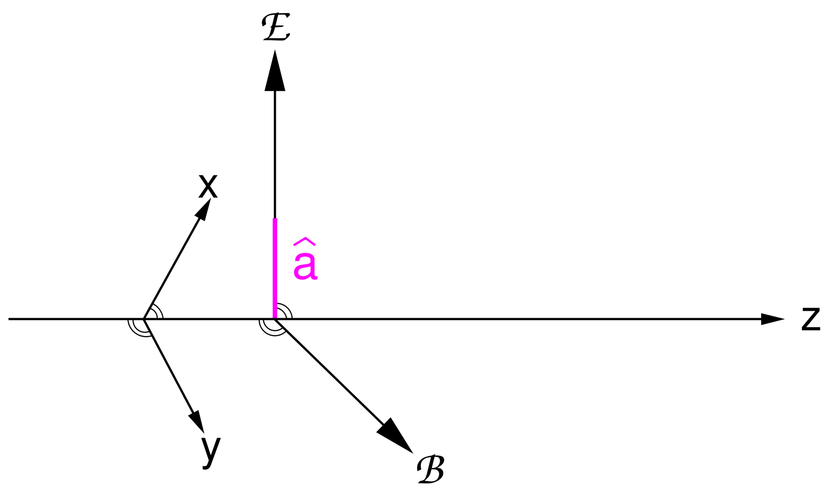


Figure C.2: Electric field and magnetic field of a linearly polarized wave. z is in the propagation direction. The plane wave front is orthogonal to z. Adapted from Burke et al., 2019

.

Since natural radio sources generally exhibit polarizations other than linear, Eq. C.2 can be expanded to the more general form with two separate complex amplitudes $\mathscr{E}_x$ and $\mathscr{E}_y$ together with the two components of $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$, where x and y then depend on z and t, as

$$\mathscr{E}(z,t) = (\hat{\mathbf{x}}\,\mathscr{E}_x + \hat{\mathbf{y}}\,\mathscr{E}_y)\, e^{i\,2\pi(\nu t - kz)} \tag{C.3}$$

with the phases entered explicitly

$$\mathscr{E}_x = |\mathscr{E}_x| e^{i\phi_x}, \qquad \mathscr{E}_y = |\mathscr{E}_y| e^{i\phi_y} \tag{C.4}$$

In the context of VGOS broadband observations, a feed horn of a single telescope receives two perpendicular linear polarization planes which represent two linearly polarized elements of the radiation. The receivers then detect the squared amplitudes $\mathscr{E}_x^2$ and $\mathscr{E}_y^2$. Following Stokes (1852), who developed the mathematical description of polarization states in the optical domain, the sum of the two outputs yields

$$\mathscr{E}_0^2 = I = \mathscr{E}_x^2(t) + \mathscr{E}_y^2(t) \tag{C.5}$$

generally called the Stokes parameter I. Since extra-galactic radio sources chosen for geodetic and astrometric VLBI initially exhibit variable states of polarization, the Stokes I_ν parameter needs to be expanded with a frequency dependent time average denoted by $\langle\rangle$ (Burke et al., 2019)

$$I_\nu = \langle \mathscr{E}_x^2(t) + \mathscr{E}_y^2(t) \rangle \tag{C.6}$$

The same also applies for the other three parameters Q, U, and V which Stokes defined to entirely describe the polarization state of an ensemble of wave trains within the emission of a broadband source

$$Q_\nu = \langle \mathscr{E}_x^2(t) - \mathscr{E}_y^2(t) \rangle \tag{C.7}$$

$$U_\nu = \langle 2\mathscr{E}_x(t)\, \mathscr{E}_y(t)\, \cos\Delta\phi \rangle \tag{C.8}$$

$$V_\nu = \langle 2\mathscr{E}_x(t)\, \mathscr{E}_y(t)\, \sin\Delta\phi \rangle \tag{C.9}$$

with $\Delta\phi = \phi_1 - \phi_2$ which is the phase difference between the two oscillating components x and y.

In radio astronomy, the components of the electric field can be represented by time-varying complex amplitudes $e_x(t)$, $e_y(t)$ (Hamaker and Bregman, 1996) leading to

$$I_\nu = \langle |e_x|^2(t) + |e_y|^2(t) \rangle \tag{C.10}$$

$$Q_\nu = \langle |e_x|^2(t) - |e_y|^2(t) \rangle \tag{C.11}$$

$$U_\nu = \langle 2|e_x|(t)\, |e_y|(t)\, \cos\Delta\phi \rangle \tag{C.12}$$

$$V_\nu = \langle 2|e_x|(t)\, |e_y|(t)\, \sin\Delta\phi \rangle \tag{C.13}$$

I is a measure of the total intensity of the wave, Q and U are the two linearly polarizend components while V is the circularly polarized component. The degrees of linear, circular and total polarization can be computed by the fractional relationships of these parameters (Thompson et al., 2007).

It should be noted that these equations only hold for single-dish cases (and/or interferometers with very short baselines). In the more general VLBI case, the Stokes parameters IQUV need to be related to visibilities (XX, XY, YX and YY) via matrices in the Radio Interferometer Measurement Equation (RIME, Hamaker et al. (1996), see below) that would account for the parallactic angles and polarization impurities of the different telescopes (among other effects). Consequently, we have to distinguish between the Stokes parameters, which are total quantities integrated over the angular extend of a source, and Stokes visibilities, which are interferometric quantities that depend on the location in the u, v plane.

C.3. Re-combination of polarized data

The (re-)combination of the observations of the two polarizations to the initial power distribution of the radio source's emission would only require forming the two combinations of $X_{(1)}X_{(2)}$ and $Y_{(1)}Y_{(2)}$. However, this would necessitate that (a) the polarization planes of the two telescopes are identical, (b) the position angle between the polarization planes of the two telescopes does not change during a scan and (c) that there are no polarization impurities. Since these conditions can hardly be fulfilled with existing equipment, other methods need to be found. The first step always is that a correlation process is set up for each of the possible combinations of $X_{(1)}X_{(2)}$, $X_{(1)}Y_{(2)}$, $Y_{(1)}X_{(2)}$, and $Y_{(1)}Y_{(2)}$. The respective cross-power spectra need to be combined before fringe fitting can be done properly. Two ways of handling the problem have been published so far. In both cases, the combination of all products is performed with the visibilities of all cross products resulting in a single array of visibilities which represent the total intensity distribution of quasi polarization-free observations.

C.3.1. Pseudo Stokes-I

In the *fourfit* program of HOPS, a linear combination of pseudo-circular parallel-hands is applied which is proportional to the Stokes I quantity (Corey 2006, Rev. 2012.01.11 and addendum). The background is that the Stokes I visibility is assumed to be similar to the (total, not interferometric) Stokes I parameter, which is the sum of intensities for two orthogonal polarizations. When, for a start, gain factors and D terms are neglected, the sum of two parallel-hands cross-correlation products should yield $X_{(1)} \cdot X_{(2)} + Y_{(1)} \cdot Y_{(2)} = I$ when the parallatic angle difference is zero.

For a more detailed explanation, the output of an imperfect linear feed is the sum of perfect and corrupted responses of the x and y planes

$$e_x = x' = x + D_x\, y \tag{C.14}$$

$$e_y = y' = y + D_y\, x \tag{C.15}$$

D represents the (complex) fractional response to the undesired polarization (D-term).

Then, the basis of the so-called Pseudo Stokes-I approach introduced by Cappallo (2016) is the construction of pseudo-circular visibilities for right-right and left-left circular polarizations (Corey, 2006) which are defined as

$$\begin{aligned}
\mathbf{RR} &= \langle (x'_{(1)}/G_{x(1)} + i y'_{(1)}/G_{y(1)})(x'_{(2)}/G_{x(2)} + i y'_{(2)}/G_{y(2)})^* \rangle \\
&= [(\mathbf{I}+\mathbf{V}) - \frac{i}{2}\mathbf{I}(-D_{x(1)} + D_{y(1)} + D^*_{x(2)} - D^*_{y(2)})]e^{-i(\psi_{(2)}-\psi_{(1)})} \qquad \text{(C.16)}\\
\mathbf{LL} &= \langle (x'_{(1)}/G_{x(1)} - i y'_{(1)}/G_{y(1)})(x'_{(2)}/G_{x(2)} - i y'_{(2)}/G_{y(2)})^* \rangle \\
&= [(\mathbf{I}-\mathbf{V}) + \frac{i}{2}\mathbf{I}(-D_{x(1)} + D_{y(1)} + D^*_{x(2)} - D^*_{y(2)})]e^{+i(\psi_{(2)}-\psi_{(1)})} \qquad \text{(C.17)}
\end{aligned}$$

The first lines of these equations represent the implicit transformation from linear to circular polarization through a complex formulation (see also Eq. C.21). The second lines follow Eq. C.26 plus complex correction terms for the respective telescope-dependent linear component D-terms. However, this is correct only if second-order products of D-terms and products of a D-term times Q, U, or V are neglected (Corey, priv. comm.). The G are the respective gains of the individual polarization planes while the $\psi_{(i)}$ are the parallactic angles (see Appendix B). * means complex conjugate.

If we now multiply the second expressions of Eqs. C.16 and C.17 with $e^{+i(\psi_{(2)}-\psi_{(1)})}$ and $e^{-i(\psi_{(2)}-\psi_{(1)})}$, respectively, and add the two equations, we find that

$$\mathbf{I} = \frac{1}{2}\left(\mathbf{RR}\, e^{+i(\psi_{(2)}-\psi_{(1)})} + \mathbf{LL}\, e^{-i(\psi_{(2)}-\psi_{(1)})}\right) \qquad \text{(C.18)}$$

In the next step, we take the first lines of Eqs. C.16 and C.17, introduce them in Eq. C.18, do the complex multiplications, and take the average gains as constants.

$$\begin{aligned}
\mathbf{I} &= \left[\frac{\langle x'_{(1)} x'^*_{(2)}\rangle}{G_{x(1)}G^*_{x(2)}} + \frac{\langle y'_{(1)} y'^*_{(2)}\rangle}{G_{y(1)}G^*_{y(2)}}\right] \cdot \cos(\psi_{(2)} - \psi_{(1)}) \qquad \text{(C.19)}\\
&+ \left[\frac{\langle x'_{(1)} y'^*_{(2)}\rangle}{G_{x(1)}G^*_{y(2)}} - \frac{\langle y'_{(1)} x'^*_{(2)}\rangle}{G_{y(1)}G^*_{x(2)}}\right] \cdot \sin(\psi_{(2)} - \psi_{(1)})
\end{aligned}$$

The factor 1/2 disappears because the cross products appear twice in the linear combination. Through the *sine* and *cosine* factors, the linear cross products are weighted relatively depending on the difference in parallactic angle (Corey, 2006).

In Cappallo (2016), the individual interferometric Pseudo Stokes-I visibility processed in the *fourfit* software is cited without the gain normalization:

$$I = (\langle X_{(1)}X_{(2)}\rangle + \langle Y_{(1)}Y_{(2)}\rangle)\cos\Delta\psi_{RS} + (\langle X_{(1)}Y_{(2)}\rangle - \langle Y_{(1)}X_{(2)}\rangle)\sin\Delta\psi_{RS} \qquad \text{(C.20)}$$

$\Delta\psi_{RS}$ is the difference in the parallactic angle (see Appendix B). The reason for the omission of the G's is that at the times of the development of the software, there was no way of estimating the gain ratios $G_{x(1)}/G_{y(1)}$ and $G_{x(2)}/G_{y(2)}$ (Corey, priv. comm.) and they had to be assumed to be unity. Any small deviation from this assumption affects the group delay only by a negligible amount except in very rare extreme cases. In contrast to this, the phase differences in the phase arguments of $G_{x(1)}/G_{y(1)}$ and $G_{x(2)}/G_{y(2)}$ are taken into account (the difference of the *fourfit* parameters pc_delay_x and pc_delay_y gives the frequency dependence, and the difference of pc_phase_offset_x and pc_phase_offset_x gives the phase offset at the *fourfit* reference frequency) (B. Corey, priv. comm.). This is done in an iterative process to determine and compensate for an y-x polarization delay and the respective phase offset[30]. Other than the standard phase calibration (phasecal) no further calibrations are applied to the visibility data before fringe fitting.

The calibrations are done for each sub-band/channel individually. Summing up the four contributions after phase calibration then provides the total intensity in the form a single set of complex visibilities for each sub-band or channel which is then used as input for the fringe fitting process (Sec. 11).

C.3.2. PolConvert

In support of observations of the Atacama Large Mm Array (ALMA), Martí-Vidal, I. et al. (2016) started to develop tools for the reduction of interferometric data from telescopes with linear and circular polarization receivers called *PolConvert* which makes use of the RIME formalisms (Hamaker et al., 1996; Smirnov, 2011). The method chosen is based on converting the linear polarization data to circular polarization data first with the electric field relationships (Martí-Vidal, I. et al., 2016)

$$e_R = \frac{1}{\sqrt{2}}(e_x - i\,e_y) \quad \text{and} \quad e_L = \frac{1}{\sqrt{2}}(e_x + i\,e_y) \tag{C.21}$$

Although this looks straight forward, phase-gain differences and amplitude-gain ratios between the two linear polarization channels of each antenna have to be applied to correct for imbalances. The numbers need to be determined beforehand with a strong calibrator source in a global cross-polarization fringe fitting algorithm which is implemented in the *PolConvert* software and which was applied to VGOS data by Alef et al. (2019).

If we assume that the voltages received at the two telescopes $v^{(1)}$ and $v^{(2)}$ are linear with respect to $\mathscr{E}$, the voltages formally refer to the electric field through

$$\mathbf{v} = \begin{bmatrix} v^{(1)} \\ v^{(2)} \end{bmatrix} = \mathbf{J} \begin{bmatrix} e^{(1)} \\ e^{(2)} \end{bmatrix} \tag{C.22}$$

J is the total *Jones* matrix comprising all propagation effects including any instrumental effects (Smirnov, 2011).

On the other hand, the four correlation products mentioned above compose the VLBI

[30]https://www.haystack.mit.edu/wp-content/uploads/2020/07/docs_hops_000_vgos-data-processing.pdf

visibility matrix **V**

$$\mathbf{V_{(1)(2)}} = \begin{bmatrix} \langle v_p^{(1)} v_p^{(2)*} \rangle & \langle v_p^{(1)} v_q^{(2)*} \rangle \\ \langle v_q^{(1)} v_p^{(2)*} \rangle & \langle v_q^{(1)} v_q^{(2)*} \rangle \end{bmatrix} \tag{C.23}$$

with p and q either being two linear (e.g., p = X and q = Y) or two circular (e.g., p = R and q = L) polarization components although mixed modes are also possible. The Fourier transform of **V** would directly be the brightness distribution matrix **B** if no source, media or instrumental polarization effects existed (**J** = **Id** = identity matrix). Since this assumption is only of hypothetical nature, calibrations for the two telescopes need to be applied so that[31]

$$\int_{\Omega} \mathbf{B}(\vec{\Omega}) \cdot exp\left[2\pi i \frac{\vec{B}\vec{\Omega}}{\lambda}\right] d\vec{\Omega} = \mathbf{J_{(1)}^{-1} V_{(1)(2)}^{LR} (J_{(2)}^{H})^{-1}} \tag{C.24}$$

Ω is the angular area of the radio source, B is the brightness distribution depending on the solid angle on the sky, while the exponential function is the kernel of the Fourier transform depending on the baseline vector $\vec{B}$ and the angular position of the source $\vec{\Omega}$. $V_{(1)(2)}^{LR}$ now is the visibilty matrix in the circular polarization domain. H represent the conjugate transpose operation, also called Hermitian transposition (Smirnov, 2011).

For a single telescope, the *Jones* matrix **J** contains the calibrations for gain **G**, D-Terms (polarization impurities) **D** and feed angle in the frame of the source **P** (Martí-Vidal et al., 2021)

$$\mathbf{J^{(i)}} = \mathbf{G^{(i)} D^{(i)} P^{(i)}} = \begin{bmatrix} G_R^{(i)} & 0 \\ 0 & G_L^{(i)} \end{bmatrix} \begin{bmatrix} 1 & D_R^{(i)} \\ D_L^{(i)} & 1 \end{bmatrix} \begin{bmatrix} e^{(i\phi_{(i)})} & 0 \\ 0 & e^{(-i\phi_{(i)})} \end{bmatrix} \tag{C.25}$$

The *Jones* matrix in Eq. C.22, which has to be composed for both telescopes together, also needs to take into account the group delay effect. Obviously, the determination of the individual contributions requires assessment of the brightness distribution matrix **B** in Eq. C.24 and is an iterative process.

The calibration of the visibility matrix is the very purpose of the *PolConvert* procedures in the context of VGOS observations. It requires the use of calibrator sources to decouple the individual effects. Furthermore, Earth rotation is needed to separate between source-related effects and those on Earth such as polarization leakage in the receiver.

Although *PolConvert* requires a more complex handling of the data than *Fourfit*, it provides advantages (I. Martí-Vidal, priv. comm.):

1. It accounts for both, phase and amplitude effects in X and Y, which improves the polarization purity of the visibilities.

2. Faraday rotation (even the one due to the Earth's magnetic field at the ionosphere) becomes a pure phase effect (which is easy to calibrate).

3. *PolConvert* provides all four Stokes parameters. The brightness distribution of all

[31] https://www.youtube.com/watch?v=J_oXou6QGpI

four Stokes parameters ($I(\alpha,\delta)$, $Q(\alpha,\delta)$, $U(\alpha,\delta)$, and $V(\alpha,\delta)$) are embedded in the brightness matrix **B** which in turn is proportional to the visibility matrix **V** with

$$\mathbf{V}_{(1)(2)} = \begin{bmatrix} \langle \mathbf{R}^{(1)}\mathbf{R}^{(2)*} \rangle & \langle \mathbf{R}^{(1)}\mathbf{L}^{(2)*} \rangle \\ \langle \mathbf{L}^{(1)}\mathbf{R}^{(2)*} \rangle & \langle \mathbf{L}^{(1)}\mathbf{L}^{(2)*} \rangle \end{bmatrix} \propto \mathbf{B} = \begin{bmatrix} \mathbf{I}+\mathbf{V} & \mathbf{Q}+i\mathbf{U} \\ \mathbf{Q}-i\mathbf{U} & \mathbf{I}-\mathbf{V} \end{bmatrix} \tag{C.26}$$

(Burke et al. (2019), see also Eq. C.24). This relationship is very useful for astronomy, in particular polarimetry, but could eventually be useful for Geodesy as well. For example, possible position changes in the source brightness peaks, due to ejections of new plasmoids in the AGN jets, which usually appear polarized, can be investigated.

Finally, after calibration, the converted circular polarized visibility data can be fed into any fringe fitting program such as *fourfit*.

D. Datum transformation parameters for the minimum case

Let us assume that we have two sets of telescope coordinates, a new one and an old (conventional) one, in which we want to transform the new one. We also assume that we do not want to question the scale parameter m between these two sets of coordinates. For this application, as a minimum we first need to identify three telescopes with good coordinates. From these, we select six components, 3 components of one telescope, 2 components of another telescope, and 1 component from yet another telescope. These six components in both frames serve as transfer information between the two frames.

To recall, the similarity transformation (Helmert transformation) excluding the scale parameter needs 6 parameters, i.e., the three translation parameters x_0, y_0, z_0, and the three rotation parameters r_x, r_y, r_z. The $\tilde{x}^{(i)}$, $\tilde{y}^{(i)}$, and $\tilde{z}^{(i)}$ ($\tilde{\mathbf{x}}$) are the coordinate components of the target data set (conventional reference frame) onto which the VLBI configuration ($x^{(i)}$, $y^{(i)}$, and $z^{(i)}$) ($\mathbf{x}$) will be mapped. The simplified Helmert transformation (cf. Eq. 13.205) without scale parameter was

$$\begin{pmatrix} \tilde{x}^{(i)} \\ \tilde{y}^{(i)} \\ \tilde{z}^{(i)} \end{pmatrix} = \begin{pmatrix} 1 & -r_z & r_y \\ r_z & 1 & -r_x \\ -r_y & r_x & 1 \end{pmatrix} \cdot \begin{pmatrix} x^{(i)} \\ y^{(i)} \\ z^{(i)} \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} \tag{D.1}$$

In the same way as in Sec. 13.8.5, we can re-order this set of equations isolating the transformation parameters. With the first three lines of Eq. D.2 for the three components of telescope (1), lines 4 and 5 of telescopes (2) and line 6 for the z component of telescope (3), we gain the observation equations for a direct solution of the problem:

$$\begin{pmatrix} \Delta x^{(1)} \\ \Delta y^{(1)} \\ \Delta z^{(1)} \\ \Delta x^{(2)} \\ \Delta y^{(2)} \\ \Delta z^{(3)} \end{pmatrix} = \begin{pmatrix} x^{(1)} - \tilde{x}^{(1)} \\ y^{(1)} - \tilde{y}^{(1)} \\ z^{(1)} - \tilde{z}^{(1)} \\ x^{(2)} - \tilde{x}^{(2)} \\ y^{(2)} - \tilde{y}^{(2)} \\ z^{(3)} - \tilde{z}^{(3)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & \tilde{z}^{(1)} & -\tilde{y}^{(1)} \\ 0 & 1 & 0 & -\tilde{z}^{(1)} & 0 & \tilde{x}^{(1)} \\ 0 & 0 & 1 & \tilde{y}^{(1)} & -\tilde{x}^{(1)} & 0 \\ 1 & 0 & 0 & 0 & \tilde{z}^{(2)} & -\tilde{y}^{(2)} \\ 0 & 1 & 0 & -\tilde{z}^{(2)} & 0 & \tilde{x}^{(2)} \\ 0 & 0 & 1 & \tilde{y}^{(3)} & -\tilde{x}^{(3)} & 0 \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ y_0 \\ z_0 \\ r_x \\ r_y \\ r_z \end{pmatrix} \tag{D.2}$$

Having determined the parameters x_0, y_0, z_0, r_x, r_y, and r_z, we can transform all coordinate components of the VLBI frame into those of the target (conventional) frame applying Eq. D.1. The determination and application of the transformation parameters to all remaining coordinate components leaves the configuration as it was initially.

In a similar way, we can treat velocity components of datum transfer telescopes. However, here the a priori information should not just be taken in the global geocentric x, y, z frame. It may be advisable to use horizontal and vertical a priori information, e.g., from a geophysical velocity model and transform this to geocentric components beforehand.

E. Derivation of NNR formulation for radio source positions

For the formulation of the NNR condition/constraint equations, we have to perform the vector product $\mathbf{k}_i \times \mathbf{\Delta k}_i$. As described in Sec. 13.9.3, the cross product in component writing with right ascensions $\tilde{\alpha}_i$ and declinations $\tilde{\delta}_i$ of the reference or target frame and their residuals or adjustments with respect to the original VLBI positions $\Delta\alpha$ and $\Delta\delta$ in angular dimensions reads

$$\mathbf{k}_i \times \mathbf{\Delta k}_i = \begin{pmatrix} \cos\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin\tilde{\delta}_i \end{pmatrix} \times \begin{pmatrix} \cos(\tilde{\alpha}_i + \Delta\alpha)\cos(\tilde{\delta}_i + \Delta\delta) - \cos\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin(\tilde{\alpha}_i + \Delta\alpha)\cos(\tilde{\delta}_i + \Delta\delta) - \sin\tilde{\alpha}_i \cos\tilde{\delta}_i \\ \sin(\tilde{\delta}_i + \Delta\delta) - \sin\tilde{\delta}_i \end{pmatrix} \tag{E.1}$$

For saving space, we now omit the indices i and, in a first step, apply trigonometric functions to the angular sums in Δk^m for the three lines m = 1, 2, 3

$$\begin{aligned} \Delta k^1 &= \cos(\tilde{\alpha} + \Delta\alpha)\cos(\tilde{\delta} + \Delta\delta) - \cos\tilde{\alpha}\cos\tilde{\delta} \\ &= (\cos\tilde{\alpha}\cos\Delta\alpha - \sin\tilde{\alpha}\sin\Delta\alpha)\cdot(\cos\tilde{\delta}\cos\Delta\delta - \sin\tilde{\delta}\sin\Delta\delta) - \cos\tilde{\alpha}\cos\tilde{\delta} \\ &= \cos\tilde{\alpha}\cos\Delta\alpha\cos\tilde{\delta}\cos\Delta\delta - \cos\tilde{\alpha}\cos\Delta\alpha\sin\tilde{\delta}\sin\Delta\delta \\ &\quad - \sin\tilde{\alpha}\sin\Delta\alpha\cos\tilde{\delta}\cos\Delta\delta + \sin\tilde{\alpha}\sin\Delta\alpha\sin\tilde{\delta}\sin\Delta\delta \\ &\quad - \cos\tilde{\alpha}\cos\tilde{\delta} \end{aligned} \tag{E.2}$$

Since the adjustments $\Delta\alpha$ and $\Delta\delta$ are small ($< 10^{-8}$), and $\Delta\alpha \cdot \Delta\delta$ is even smaller, we can first replace $\sin\alpha/\delta = \alpha/\delta$ and $\cos\alpha/\delta = 1$ and then sum up all elements to

$$\begin{aligned} \Delta k^1 &= -\cos\tilde{\alpha}\sin\tilde{\delta}\Delta\delta - \sin\tilde{\alpha}\cos\tilde{\delta}\Delta\alpha + \sin\tilde{\alpha}\sin\tilde{\delta}\Delta\alpha\Delta\delta \\ &= -\cos\tilde{\alpha}\sin\tilde{\delta}\Delta\delta - \sin\tilde{\alpha}\cos\tilde{\delta}\Delta\alpha \end{aligned} \tag{E.3}$$

Then we do the same for the other two lines:

$$\begin{aligned} \Delta k^2 &= \sin(\tilde{\alpha} + \Delta\alpha)\cos(\tilde{\delta} + \Delta\delta) - \sin\tilde{\alpha}\cos\tilde{\delta} \\ &= (\sin\tilde{\alpha}\cos\Delta\alpha + \cos\tilde{\alpha}\sin\Delta\alpha)\cdot(\cos\tilde{\delta}\cos\Delta\delta - \sin\tilde{\delta}\sin\Delta\delta) - \sin\tilde{\alpha}\cos\tilde{\delta} \\ &= \sin\tilde{\alpha}\cos\Delta\alpha\cos\tilde{\delta}\cos\Delta\delta - \sin\tilde{\alpha}\cos\Delta\alpha\sin\tilde{\delta}\sin\Delta\delta \\ &\quad + \cos\tilde{\alpha}\sin\Delta\alpha\cos\tilde{\delta}\cos\Delta\delta - \cos\tilde{\alpha}\sin\Delta\alpha\sin\tilde{\delta}\sin\Delta\delta \\ &\quad - \sin\tilde{\alpha}\cos\tilde{\delta} \\ &= -\sin\tilde{\alpha}\sin\tilde{\delta}\Delta\delta + \cos\tilde{\alpha}\cos\tilde{\delta}\Delta\alpha - \cos\tilde{\alpha}\sin\tilde{\delta}\Delta\alpha\Delta\delta \\ &= -\sin\tilde{\alpha}\sin\tilde{\delta}\Delta\delta + \cos\tilde{\alpha}\cos\tilde{\delta}\Delta\alpha \end{aligned} \tag{E.4}$$

$$\begin{aligned} \Delta k^3 &= \sin(\tilde{\delta} + \Delta\delta) - \sin\tilde{\delta} \\ &= \sin\tilde{\delta}\cos\Delta\delta + \cos\tilde{\delta}\sin\Delta\delta - \sin\tilde{\delta} \\ &= \cos\tilde{\delta}\Delta\delta \end{aligned} \tag{E.5}$$

Next comes the vector product with the pre-processed $\mathbf{\Delta k}$ vector elements

$$\mathbf{k} \times \mathbf{\Delta k} = \begin{pmatrix} \cos\tilde{\alpha}\cos\tilde{\delta} \\ \sin\tilde{\alpha}\cos\tilde{\delta} \\ \sin\tilde{\delta} \end{pmatrix} \times \begin{pmatrix} -\cos\tilde{\alpha}\sin\tilde{\delta}\Delta\delta - \sin\tilde{\alpha}\cos\tilde{\delta}\Delta\alpha \\ -\sin\tilde{\alpha}\sin\tilde{\delta}\Delta\delta + \cos\tilde{\alpha}\cos\tilde{\delta}\Delta\alpha \\ \cos\tilde{\delta}\Delta\delta \end{pmatrix} \tag{E.6}$$

$$\begin{aligned} (k \times \Delta k)^1 &= \sin\tilde{\alpha}\cos\tilde{\delta}\cos\tilde{\delta}\Delta\delta - \sin\tilde{\delta}(-\sin\tilde{\alpha}\sin\tilde{\delta}\Delta\delta + \cos\tilde{\alpha}\cos\tilde{\delta}\Delta\alpha) \\ &= \sin\tilde{\alpha}\cos^2\tilde{\delta}\Delta\delta + \sin\tilde{\alpha}\sin^2\tilde{\delta}\Delta\delta - \cos\tilde{\alpha}\sin\tilde{\delta}\cos\tilde{\delta}\Delta\alpha) \\ &= (\cos^2\tilde{\delta} + \sin^2\tilde{\delta})\sin\tilde{\alpha}\Delta\delta - \cos\tilde{\alpha}\sin\tilde{\delta}\cos\tilde{\delta}\Delta\alpha \\ &= \sin\tilde{\alpha}\Delta\delta - \cos\tilde{\alpha}\sin\tilde{\delta}\cos\tilde{\delta}\Delta\alpha \end{aligned} \tag{E.7}$$

$$\begin{aligned} (k \times \Delta k)^2 &= \sin\tilde{\delta}(-\cos\tilde{\alpha}\sin\tilde{\delta}\Delta\delta - \sin\tilde{\alpha}\cos\tilde{\delta}\Delta\alpha) - \cos\tilde{\alpha}\cos\tilde{\delta}\cos\tilde{\delta}\Delta\delta \\ &= -\sin\tilde{\alpha}\sin\tilde{\delta}\cos\tilde{\delta}\Delta\alpha - \cos\tilde{\alpha}\sin^2\tilde{\delta}\Delta\delta - \cos\tilde{\alpha}\cos^2\tilde{\delta}\Delta\delta \\ &= -\sin\tilde{\alpha}\sin\tilde{\delta}\cos\tilde{\delta}\Delta\alpha - (\sin^2\tilde{\delta} + \cos^2\tilde{\delta})\cos\tilde{\alpha}\Delta\delta \\ &= -\sin\tilde{\alpha}\sin\tilde{\delta}\cos\tilde{\delta}\Delta\alpha - \cos\tilde{\alpha}\Delta\delta \end{aligned} \tag{E.8}$$

$$\begin{aligned} (k \times \Delta k)^3 &= \cos\tilde{\alpha}\cos\tilde{\delta}(-\sin\tilde{\alpha}\sin\tilde{\delta}\Delta\delta + \cos\tilde{\alpha}\cos\tilde{\delta}\Delta\alpha) \\ &\quad - \sin\tilde{\alpha}\cos\tilde{\delta}(-\cos\tilde{\alpha}\sin\tilde{\delta}\Delta\delta - \sin\tilde{\alpha}\cos\tilde{\delta}\Delta\alpha) \\ &= -\sin\tilde{\alpha}\cos\tilde{\alpha}\sin\tilde{\delta}\cos\tilde{\delta}\Delta\delta + \cos^2\tilde{\alpha}\cos^2\tilde{\delta}\Delta\alpha + \sin^2\tilde{\alpha}\cos^2\tilde{\delta}\Delta\alpha + \sin\tilde{\alpha}\cos\tilde{\alpha}\sin\tilde{\delta}\cos\tilde{\delta}\Delta\delta \\ &= (\sin^2\tilde{\alpha} + \cos^2\tilde{\alpha})\cos^2\tilde{\delta}\Delta\alpha \\ &= \cos^2\tilde{\delta}\Delta\alpha \end{aligned} \tag{E.9}$$

F. Some explanations of *fourfit* fringe fitting plots

This is a partly complete description of what is displayed in a plot which is provided by the HOPS program *fourfit* for each observation for diagnostic purposes. The contents and order of sub-plots and tables originates from the original ASCII character plot emulation of the FRNGE fringe fitting program of the 1980ies by Alan Rogers, Haytsack Observatory. The entire plot for an S/X/K single band observation looks as below. For VGOS observations, there are a few additions (Sec. F.3).

Figure F.1: Complete *fourfit* fringe plot of X band observation of baseline HART15M - WETTZELL

F.1. Individual fringe plots

Header of the *fourfit* fringe plot

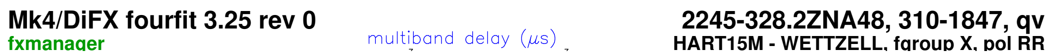


Figure F.2: Header information.

Fig. F.2: Left, *fourfit* revision number, right, 8-character source name (2245-328) separated by "." from encrypted fringe fit time, scan name as day of year-hhmm (310-1847), 2*1-character telescope code (q and v) to be used as single character telescope identifiers everywhere below, baseline in readable names, frequency band (X), polarizations of both telescopes (R RCP, X linear polarization, Y linear polarization).

Main multiband delay and delay rate plot

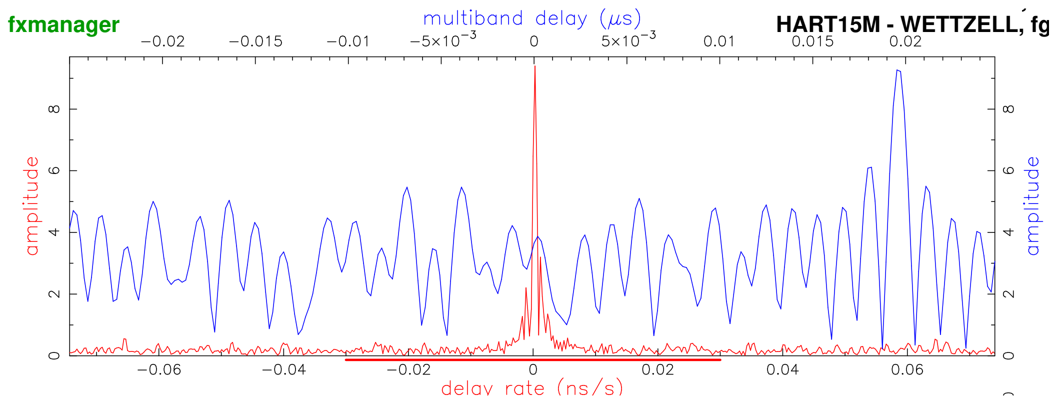
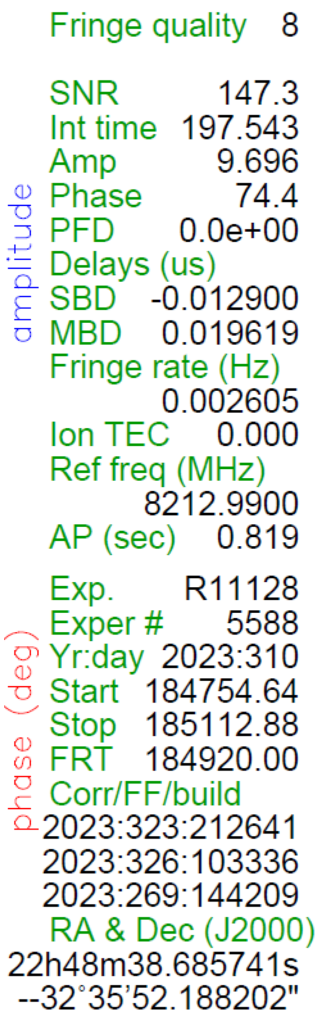


Figure F.3: Multiband delay and delay rate.

Fig. F.3: Multiband delay resolution function (blue) superimposed by delay rate spectrum (red). Delay in μs. Delay window is one ambiguity spacing (here 50 ns). Delay rate (DR) is depicted in ns/s. The DR window is $1/(2 \cdot N_{ap})/\nu$ where N_{ap} is number of APs. Both peaks should be centered to indicate proper a priori modeling. Centered peak of delay rate, i.e., delay rate near 0 ns/s, is critical for optimum SNR.

Right hand listing of *fourfit* fringe plot



Explanations:

- Fringe quality
  Quality codes (QC) 9,8,7,6,5,4,3,2,1 = no error condition, 9 is best, reductions for increased scatter in phases and/or amplitudes.

  = 0 Fringes not detected,
  Probability of false detection (PFD) > 1.e-4.
  = B Interpolation error in fourfit.
  = D No data in one or more frequency channels.
  = E Max fringe amplitude at the edge of SBD, MBD or DR window.
  = F "Fork" problem in processing.
  = G Fringe amp. in one or more channels is < 0.5 mean amp. (for SNR > 20).
  = H Low pcal-amplitude.
  = N No valid correlator data.

- SNR = Signal-to-noise ratio
- Int time = Integration time in seconds
- Amp = correlation amplitude in units of 100 ppm
- Phase = residual fringe phase
- PFD = probability of false detection, if SNR is small (< 10)
- SBD = residual singleband delay in μs
- MBD = residual multiband delay in μs
- Fringe rate in Hz, $\dot{f} = \dot{\tau} \cdot \nu$
- Ion TEC = Total ionospheric electron content in TEC units (VGOS only)
- Ref freq = Reference frequency in MHz
- AP = length of accumulation period of correlation in seconds
- Exp. = experiment name in IVS master schedule
- Exper # = experiment number of correlator center
- Yr:day = Year and day of session
- Start = start time of scan/observation in hhmmss.ss
- Stop = stop time of scan/observation in hhmmss.ss
- FRT = Fringe reference epoch of scan/observation in hhmmss.ss as selected by *fourfit* close to middle of scan (mod 10 s)
- Corr/FF/build = Correlator, *fourfit* versions
- RA & Dec (J2000) = position of target in J2000.0

Cross power spectrum and singleband delay

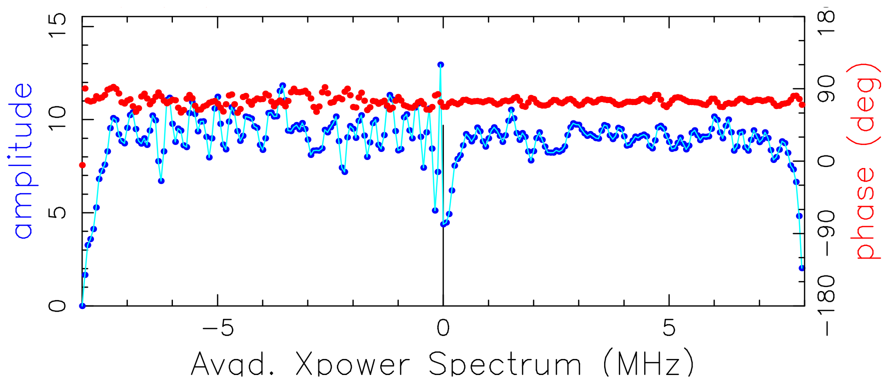


Figure F.4: Cross power spectrum.

Fig. F.4: Cross power spectrum of the two lower and eight upper side band sub-bands/channels of the current X band frequency setup in amplitude (blue) and phase (red). The plot displays the respective **averages** of all sub-bands/channels after phase calibration was applied. The drop in amplitudes at the band edges indicate the bandpass shape of one or both telescopes.

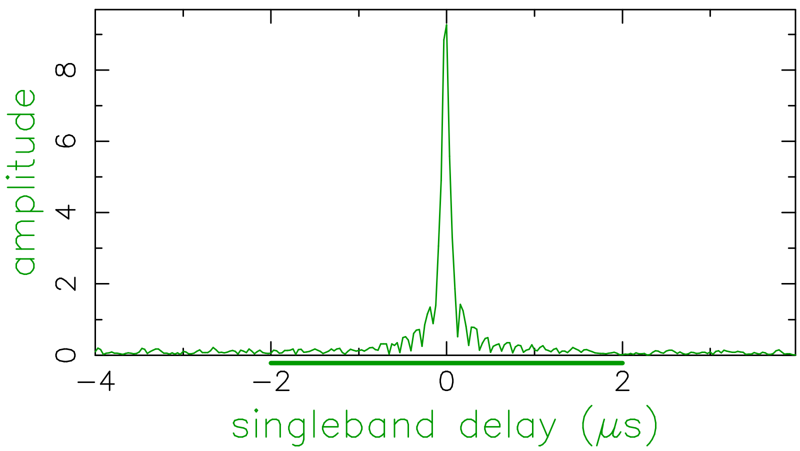


Figure F.5: Single band delay plot.

Fig. F.5: The singleband delay plot depicts the shape of the **average** correlation function as unitless amplitude versus delay in μs. The average is formed (incoherently) over all sub-bands correlated and employed in the fringe fit.

Amplitudes and phases versus time per sub-band/channel

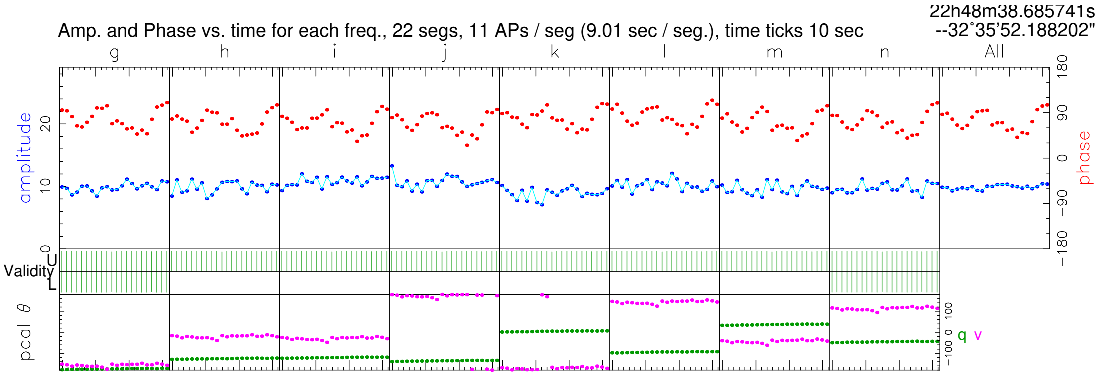


Figure F.6: Amplitudes and phases versus time for each sub-band/channel.

Fig. F.6: Time series of averaged amplitudes and phases. Sub-bands/channels are normally named a-f for S band and g-n for X band. Averaging is done over a time segment of specified length depending on the space available for plotting (here 9.01 s). Details of time scaling and averaging are depicted in top line.

- Top section: amplitudes in blue and phases in red versus time after phase calibration and fringe fit for each sub-band/channel.
- Middle section: data validity, first line for USB and second line for LSB.
- Third section: phase calibration phases in degrees versus time. Green color for reference telescope, magenta for remote telescope. Averaging is first done for N_{ap} APs as defined in "Pcal mode" (Sec. 11.4 and Fig. F.11) and then for proper plotting. If N_{ap} is set to include all APs, as can be done with command 'pc_period 9999' in the *fourfit* control file or was done with the obsolete option "normal", the dots will form a flat line.

All time series should be flat with no systematics and little noise.

F.2. Numerical information

Numerical information section 1

	8212.99	8252.99	8352.99	8512.99	8732.99	8852.99	8912.99	8932.99	Freq (MHz)	All
	78.1	71.7	70.0	68.5	79.3	83.7	71.0	72.1	Phase	74.4
	9.3	9.4	10.3	10.1	8.5	9.8	9.1	9.2	Ampl.	9.4
	128.6	128.5	128.8	128.7	128.4	128.5	128.4	128.6	Sbd box	128.6
U/L	242/242	242/0	242/0	242/0	242/0	242/0	242/0	242/242	APs used	
q	493.4	494.7	498.6	493.7	-495.9	-488.9	-494.9	-494.3	PC R delays (ns)	
v	-272.8	-272.2	-272.1	-274.9	-258.0	-258.8	-258.2	-262.7	PC R delays (ns)	
q:v	-169:-156	-125:-21	-120:-29	-158:175	4:-172	-86:143	42:-43	-46:113	PC phase	
q:v	-5:0	-1:0	-1:0	23:0	1:0	-8:0	-6:0	1:0	Manl PC	
q	28	25	23	22	24	11	15	16	PC amp	
v	60	54	63	60	59	64	61	59		

Figure F.7: Numerical information section 1 of fringe plot.

- Line 1 (Freq (MHz)): Frequency sub-bands employed in fringe fit.
- Line 2 (Phase): Estimated fringe phases for all frequency sub-bands employed in fringe fit, and average phase (last column)
- Line 3 (Ampl): Fringe amplitude for all frequency sub-bands employed in fringe fit, and average amplitude (last column)
- Line 4 (Sbd box): Singleband delay box. Encrypted singleband delays for all frequency sub-bands employed in fringe fit, and average (last column). The Sbd box is computed with
$$\text{Sbd box} = 2 \cdot \frac{\text{SBD}}{\text{Tsamp}} + \text{Nlag} + 1 \tag{F.1}$$
where Tsamp is the interval between data samples to be computed with 1/Sample rate(MSamp/s), in this case TSamp = 62.5 ns. Nlag is the number of time lags in cross-correlation, listed in the second block of numerical information. Nlag can also be calculated from the total width of the green "singleband delay" function plot by Nlag = SBD_plot_width / Tsamp.
At the advent of fringe plots, the multiband delay ambiguity spacing was called a "box" to emphasize that it is an integer. So, the multiband and singleband delays could easily be compared after correcting for an integer number of boxes (A. Rogers, priv. comm.).
- Line 5 (APs used): Accumulation periods (N_{ap}) used in fringe fit for all frequency sub-bands employed in fringe fit separated for upper and lower sideband data by /
- Line 6/7 (PC R delays (ns)): Phase calibration delays in ns, averaged over all APs, R for RCP, for both telescopes underneath, ambiguous by 1 μs due to 1 MHz tone separation (This is the reason why sub-bands k-n apparently have opposite signs of sub-bands g-j. In fact, they only differ by ≈10 ns.).
- Line 8 (PC phase): Phase calibration phases in degrees, averaged over all APs, as derived from phase calibration delays for both telescopes together (here q and v) separated by : (Phase calibration uses the individual phasecal phases calculated for each of the APs.).
- Line 9 (Manl PC): Manual phase calibration phases in degrees for both telescopes, separated by :. Optional: Additive phase offsets for regular phase calibration phases.
- Line 10/11 (PC Amp): Phase calibration amplitudes for both telescopes underneath in units of 0.1% of signal voltage. If phasecal signal level is correct and phase calibration

is applied, the color is green. If phasecal signal level is too small or too high, the color is red. It is also red if there is no phasecal data available. Then this is also indicated by "1000" as in Fig. F.8.

Manual phase calibration info

K:v	0:-18	0:116	0:105	0:-53	0:-41	0:-89	0:83	0:-121	PC phase
K:v	153:0	-77:0	-50:0	-309:0	-28:0	-89:0	-38:0	-10:0	Manl PC
K	1000	1000	1000	1000	1000	1000	1000	1000	PC amp
v	79	74	84	82	79	85	81	79	

Figure F.8: Phase calibration amplitudes II.

Fig. F.8: If phase calibration signal is missing entirely, the respective telescope line(s) are depicted in red, mostly with indicator "1000".

Sub-band/channel identifiers

q	X06UR,X06LR	X07UR	X08UR	X09UR	X10UR	X11UR	X12UR	X13UR,X13LR	Chan ids Tracks
v	X06UR,X06LR	X07UR	X08UR	X09UR	X10UR	X11UR	X12UR	X13UR,X13LR	Chan ids Tracks

Figure F.9: Sub-band/channel identifiers.

Fig. F.9: Each sub-band is assigned a unique channel identifier. X06UR stands for X band, channel ID #6, upper sideband, right circular polarization. Channel ID counter normally starts with first S band sub-band. The "Tracks" lines are remnants from old tape drive operations.

Numerical information section 2

Group delay (usec) (MODEL)	1.75147690684E+04	Apriori delay (usec)	1.75147494491E+04	Resid mbdelay (usec)	1.96193E-02	+/-	3.9E-06
Sband delay (usec)	1.75147365491E+04	Apriori clock (usec)	-1.3893859E+01	Resid sbdelay (usec)	-1.29000E-02	+/-	3.3E-04
Phase delay (usec)	1.75147494742E+04	Apriori clockrate (us/s)	-4.3579246E-07	Resid phdelay (usec)	2.51534E-05	+/-	1.9E-07
Delay rate (us/s)	-2.93413564764E-01	Apriori rate (us/s)	-2.93413881998E-01	Resid rate (us/s)	3.17234E-07	+/-	2.3E-09
Total phase (deg)	102.3	Apriori accel (us/s/s)	-2.11116051767E-05	Resid phase (deg)	74.4	+/-	0.6

Figure F.10: Total, a priori and residual observables.

Fig. F.10: The correlation and fringe fitting processes only determine residual quantities relative to a priori model values. The total quantities are composed in this text block. N.B.: Here, the totals are geocentric delays and phases. In parentheses behind the "Group delay (usec)", the setting of the multiband delay reference in the *fourfit* control file (mbd_anchor) is documented. It controls what the reference for the ambiguity-affected multiband delay is. This can either be "MODEL", the correlator model delay, or "sbd", the singleband delay (Sec.12.4).

Numerical information section 3.

```
               RMS    Theor.    Amplitude          9.696 +/- 0.066      Pcal mode: MULTITONE,  MULTITONE  PC period (AP's) 5,  5
ph/seg (deg)   18.2   1.8       Search (512X256)   9.073                Pcal rate: 0.000E+00,  0.000E+00  (us/s)                    sb window (us)    -2.000   2.00
amp/seg (%)    4.1    3.2       Interp.            0.000                Bits/sample: 2x2          SampCntNorm: disabled             mb window (us)    -0.025   0.02
ph/frq (deg)   5.8    1.1       Inc. seg. avg.     9.895                Data rate(MSamp/s): 16 MBpts 256 Amb 0.050 us               dr window (ns/s)  -0.030   0.03
amp/frq (%)    5.9    1.9       Inc. frq. avg.     9.432                Data rate(Mb/s): 320      nlags: 128   t_cohere infinite    ion window (TEC)   0.00    0.00
q: az 235.8 el 77.3 pa 60.5          v: az 178.0 el 8.4 pa -1.3          u,v (fr/asec) 186.513  -746.428                                simultaneous interpolator
Control file: cf_5588    Input file: /Exps/r11128/v1/5588/310-1847/qv..2ZNA48    Output file: Suppressed by test mode
```

Figure F.11: Additional information.

Fig. F.11:

- Keyword "Pcal mode" indicates phase calibration mode. "MULTITONE, MULTITONE PC period (AP's) 5, 5" means that phase calibration phases are computed from multiple tone phases (Sec. 11.4) averaged over 5 accumulation periods for both the reference and the remote telescope.
- "Bits/sample" indicates sampling scheme, here 2x2 means 2 bit quantization twice for every Hz in baseband frequency.
- "Sample rate" in MSamples/s is listed for one sub-band/channel.
- "Data rate" in Mb/s is sum over all sub-bands/channels.
- "nlags" are number of frequency lags employed in first Fourier transformation of FX correlator.

F.3. VGOS fringe plot

In fringe plots of VGOS observations, additional information is displayed such as the dTEC estimation function superimposed on the single band delay. The cross power spectrum has only lower side band frequencies because all sub-bands are LSB.

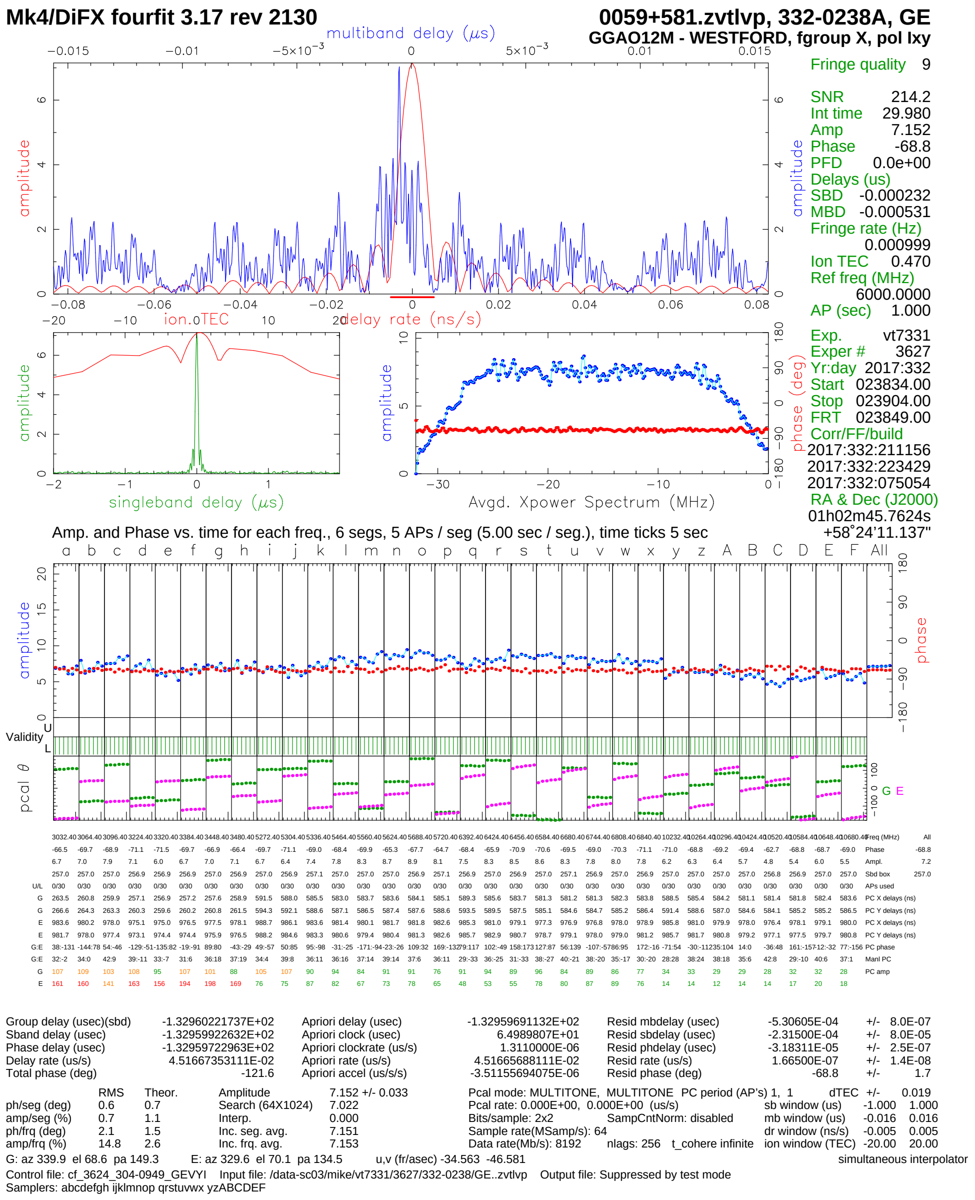


Figure F.12: Complete *fourfit* fringe plot of VGOS observation.

Special items of VGOS *fourfit* fringe plots.

- Phase calibration phases are normally computed for each accumulation period separately and independently of other ACs.
- Ionospheric dTEC estimation function is superimposed on singleband delay plot (in red).
- Cross power spectrum only shows lower sideband.
- Phase calibration delays are listed for both, X and Y, polarizations.
- There are no lines for sub-band/channel identifiers.
- Phase calibration amplitudes in Fig. F.8 are coded in yellow if phasecal amplitude is only slightly above or below the 1 - 10% (10 - 100) specification, i.e., if 10.1% > amp > 14.9% or if 0.9% > amp > 0.4%.
- Sub-band/channel denominators range from a-F

G. Geometric delays - Some rules of thump

According to Eq. 2.7, the geometric time delay in a VLBI triangle is

$$\tau = -\frac{1}{c}\mathbf{b} \cdot \mathbf{k} \tag{G.1}$$

with $\mathbf{b}$ the 3D baseline vector, $\mathbf{k}$ the unit vector in source direction, and c the velocity of light. For the purpose of rule-of-thump calculations, we can simplify this equation by just considering the 2D case, which reads

$$\tau = -\frac{1}{c} \cdot b \cdot \cos\theta \tag{G.2}$$

where b is the length of the baseline and θ is the angle between source direction and baseline vector (Fig. G.1).

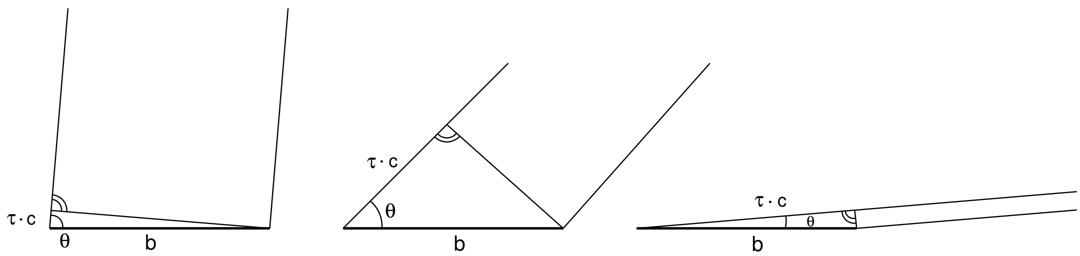


Figure G.1: Examples of VLBI triangles. Left (a) with (close to) maximum angle of $\theta = 90°$, middle (b) with angle $\theta = 45°$, and right (c) with (close to) minimum angle $\theta = 0°$.

With Eq. G.2 we can compute a little table of rule-of-thump delays (Tab. G.1), which may be helpful for ballpoint estimates. Note that for these delays, the reference telescope is the left one of Fig. G.1.

Table G.1: Delays depending on baseline lengths and angle θ.

	$\theta = 0°$	15°	30°	45°	60°	75°	90°
1000 km	-3.3 ms	-3.2 ms	-2.9 ms	-2.4 ms	-1.7 ms	-0.9 ms	0.0 ms
2000 km	-6.7 ms	-6.4 ms	-5.8 ms	-4.7 ms	-3.3 ms	-1.7 ms	0.0 ms
3000 km	-10.0 ms	-9.7 ms	-8.7 ms	-7.1 ms	-5.0 ms	-2.6 ms	0.0 ms
4000 km	-13.3 ms	-12.9 ms	-11.5 ms	-9.4 ms	-6.7 ms	-3.5 ms	0.0 ms
5000 km	-16.7 ms	-16.1 ms	-14.4 ms	-11.8 ms	-8.3 ms	-4.3 ms	0.0 ms
6000 km	-20.0 ms	-19.3 ms	-17.3 ms	-14.1 ms	-10.0 ms	-5.2 ms	0.0 ms
7000 km	-23.3 ms	-22.5 ms	-20.2 ms	-16.5 ms	-11.7 ms	-6.0 ms	0.0 ms
8000 km	-26.7 ms	-25.8 ms	-23.1 ms	-18.9 ms	-13.3 ms	-6.9 ms	0.0 ms
9000 km	-30.0 ms	-29.0 ms	-26.0 ms	-21.2 ms	-15.0 ms	-7.8 ms	0.0 ms
10000 km	-33.3 ms	-32.2 ms	-28.9 ms	-23.6 ms	-16.7 ms	-8.6 ms	0.0 ms

The delay is always zero if the source vector is perpendicualr to the baseline. For very small angles θ, the delay is almost as large as the baseline itself divided by c. However, even more interesting are rule-of-thump estimates of sensitivities of the delays for changes in θ. These can originate, for example, from shifts in the position of the radio source observed or from variations in Earth orientation, which the baseline responds to in its orientation. For this purpose, we build the derivative of Eq. G.2 with respect to θ and multiply it with the respective shift $\Delta\theta$ yielding

$$\Delta\tau = \frac{d\tau}{d\theta} \cdot \Delta\theta = \frac{1}{c} \cdot b \cdot \sin\theta \, \Delta\theta. \tag{G.3}$$

Assuming a variation in the relative orientation of 1 mas, we find the following table of responses in the observed delay:

Table G.2: Sensitivity of delay to 1 mas variation in θ depending on baseline lengths and angle θ.

	$\theta = 0°$	15°	30°	45°	60°	75°	90°
1000 km	0.0 ps	4.2 ps	8.1 ps	11.4 ps	14.0 ps	15.6 ps	16.2 ps
2000 km	0.0 ps	8.4 ps	16.2 ps	22.9 ps	28.0 ps	31.2 ps	32.3 ps
3000 km	0.0 ps	12.5 ps	24.2 ps	34.3 ps	42.0 ps	46.8 ps	48.5 ps
4000 km	0.0 ps	16.7 ps	32.3 ps	45.7 ps	56.0 ps	62.4 ps	64.6 ps
5000 km	0.0 ps	20.9 ps	40.4 ps	57.1 ps	70.0 ps	78.0 ps	80.8 ps
6000 km	0.0 ps	25.1 ps	48.5 ps	68.6 ps	84.0 ps	93.6 ps	97.0 ps
7000 km	0.0 ps	29.3 ps	56.6 ps	80.0 ps	98.0 ps	109 ps	113 ps
8000 km	0.0 ps	33.5 ps	64.6 ps	91.4 ps	112 ps	125 ps	129 ps
9000 km	0.0 ps	37.6 ps	72.7 ps	103 ps	126 ps	140 ps	145 ps
10000 km	0.0 ps	41.8 ps	80.8 ps	114 ps	140 ps	156 ps	162 ps

Tab. G.2 shows that observations in the direction of the baseline are insensitive to variations in the orientation. On the other hand, observations to radio sources perpendicular to the baseline are affected most by variations in the relative orientation of baseline and unit vector in source direction.

H. Axis Offset Determination

As we have seen in Sec. 5.1.1, the knowledge of the axis offset (AO) is of paramount importance for reliable terrestrial reference frame determinations with VLBI. One way of determination is the calculation from local survey data. The easiest concept for an azimuth-elevation telescope is the following. After the two end points of the elevation axis, L(eft) and R(ight) as seen in the direction of the radio source, have been visualized by markers, their positions in a local frame can be measured by simple forward intersects for a series of telescope pointing azimuths. This means that after the first pair of measurements has been performed, the telescope is rotated (e.g., clockwise) by a certain increment in azimuth, e.g., 15°. Then the forward intersect is repeated for the two end points for all other azimuth angles α_i of a full circle. Fig. H.1 shows the situation for $\alpha = 0°$.

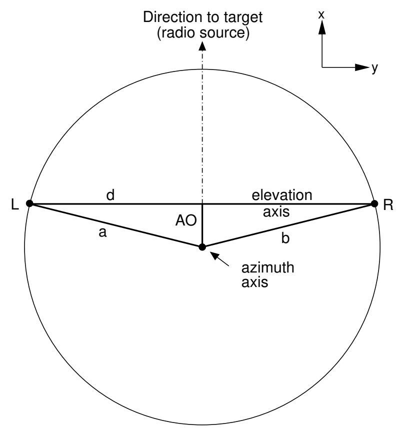


Figure H.1: Azimuth and elevation axis in the plane of the elevation axis with two end points L(eft) and R(ight) with positive axis offset AO.

The targets **L** and **R** describe two circles with same center point but with different mean radii $\bar{a}$ and $\bar{b}$. The center points can be estimated by least squares adjustments and the average of the two center points is the position of the azimuth axis in the plane of the elevation axis (X_O, Y_O). The level of agreement of the two estimates is a good indicator of the quality of the local forward intersects performed in the survey. Even more conclusive are the (separate!) series of radii a_i and b_i with respect to the telescope's azimuth directions. Most probably, the means will not be identical due to telescope construction properties. The noise again reflects the quality of the measurements.

The necessary computations of the axis offset o is repeated for each azimuth angle position of the radio telescope and consists of the following steps.

1. Compute a(α_i) and b(α_i) with the mean coordinates of the position of the azimuth axis by simple Pythagorean theorem.

2. Compute d(α_i) from the coordinates of the two end points L and R by Pythagorean theorem.

3. Compute AO_i with

$$AO_i = \sqrt{a_i^2 - \left(\frac{d_i^2 - a_i^2 + b_i^2}{2d_i}\right)^2} \tag{H.1}$$

The derivation of this formula is presented at the end of this section.

The mean of all AO_i is the absolute value of the axis offset but beware that there still is the sign of AO! The reason is that the elevation axis may well lie behind the azimuth axis as depicted in Fig. H.2.

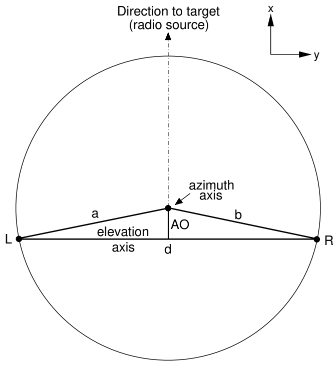


Figure H.2: Azimuth and elevation axis in the plane of the elevation axis with two end points L(eft) and R(ight) with negative axis offset AO.

An easy way to figure out whether the elevation axis lies in front of the azimuth axis and the axis offset is positive, is by computing the respective azimuths of **a** and **b** for each telescope azimuth. The azimuth from the center, i.e., the azimuth axis with coordinates X_0, Y_0 to, e.g., target L with coordinates X_L, Y_L is computed with

$$\beta_{L(\alpha_i)} = \arctan\left(\frac{Y_L - Y_O}{X_L - X_O}\right) \tag{H.2}$$

Please be aware that the X coordinate component normally points North and the Y component points East following common geodetic practice. For $\beta_{R(\alpha_i)}$ just replace the respective components. Then assuming L and R as shown in Fig. H.1 and H.2, the difference between $\beta_{L(\alpha_i)}$ and $\beta_{R(\alpha_i)}$ in the sense

$$\Delta\beta_i = \beta_{R(\alpha_i)} - \beta_{L(\alpha_i)} \tag{H.3}$$

is either smaller than 180° or larger. In the first case ($\Delta\beta_i < 180°$), the axis offset is positive, in the second case ($\Delta\beta_i > 180°$), it is negative.

This should be checked carefully for all telescope positions α_i because there certainly are variations in $\Delta\beta_i$. The smaller AO, the higher the probability for toggling signs. Then, a plausible decision has to be taken on the basis of the data and it may even be necessary to decide that $AO = 0$.

Derivation of Eq. H.1

The derivation of Eq. H.1 as denominated according to Fig. H.3 is based on two rectangular triangles with o in common and the p/q formula by writing

$$o^2 = a^2 - p^2 = b^2 - q^2. \tag{H.4}$$

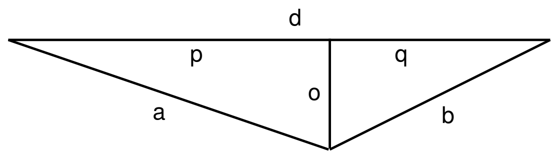


Figure H.3: Triangle for derivation of Eq. H.1.

Simple arithmetic produces

$$\frac{p+q}{2} = \frac{d}{2} = \frac{d^2}{2d} \tag{H.5}$$

and

$$\frac{p-q}{2} = \frac{(b+a)(b-a)}{2d} = \frac{b^2-a^2}{2d} \tag{H.6}$$

Adding Eq. H.5 and H.6 provides

$$\frac{p+q}{2} + \frac{p-q}{2} = p = \frac{d^2}{2d} + \frac{b^2-a^2}{2d} = \frac{d^2-a^2+b^2}{2d} \tag{H.7}$$

Then we insert Eq. H.7 in Eq. H.4 for p and end up with Eq. H.1.

Some more comments

The coordinates of targets L and R can of course also be determined by indirect methods, for example, if multiple targets are placed on the moving structure arbitrarily. Then, normally cylinders are constructed mathematically and the vector between L and R can be computed purely mathematically.

Finally, it should also be mentioned that axis offsets have been estimated in Level-2 data analysis (Krásná et al., 2014; Kurdubov and Skurikhina, 2010; Nilsson et al., 2017). However, often estimates and local surveys disagree by up to a centimeter for unknown reasons (Conference slides by Nilsson et al. (2019))[32]. Speculation is that the estimates also absorb other effects with similar signature. For this reason, local surveys are always to be preferred.

[32]https://www.oan.es/evga2019/PDF/A2/A2-3-EVGA2019.Nilsson.pdf

References

Abbondanza, C., & Sarti, P. (2010). Effects of illumination functions on the computation of gravity-dependent signal path variation models in primary focus and Cassegrainian VLBI telescopes. *J Geod*, *84*(8), 515–525. https://doi.org/10.1007/s00190-010-0389-z

Alef, W., Anderson, J. M., Bernhart, S., de Vicente, P., González García, J., Haas, R., La Porta, L., Martí Vidal, I., Müskens, A., Nothnagel, A., Plötz, C., Rottmann, H., Savolainen, T., Schüler, T., Small, D., & Wagner, J. (2019). The EUropean-VGOS Project. In R. Haas, S. Garcia-Espada, & J. A. López Fernández (Eds.), *Proceedings of the 24th European VLBI Group for Geodesy and Astrometry Working Meeting* (pp. 107–111).

Alizadeh, M. M., Wijaya, D. D., Hobiger, T., Weber, R., & Schuh, H. (2013). Ionospheric effects on microwave signals. In J. Böhm & H. Schuh (Eds.), *Atmospheric effects in space geodesy* (pp. 35–71). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-36932-2_2

Altamimi, Z., Rebischung, P., Métivier, L., & Collilieux, X. (2016). ITRF2014: A new release of the international terrestrial reference frame modeling nonlinear station motions. *Journal of Geophysical Research: Solid Earth*, *121*(8), 6109–6131. https://doi.org/10.1002/2016jb013098

Amiri-Simkooei, A. R., ASCE, M., Asgari, J., Zangeneh-Nejad, F., & Zaminpardaz, S. (2012). Basic concepts of optimization and design of geodetic networks. *J. Surv. Eng.*, *138*(4), 172–183.

Anderson, J. M., & Xu, M. H. (2018). Source Structure and Measurement Noise Are as Important as All Other Residual Sources in Geodetic VLBI Combined [doi:10.1029/2018JB015550]. *Journal of Geophysical Research (Solid Earth)*, *123*(11), 10, 162–10, 190. https://doi.org/10.1029/2018JB015550

Angermann, D., Drewes, H., Krügel, M., Meisel, B., Gerstl, M., Kelm, R., Müller, H., Seemüller, W., & Tesmer, V. (2004). ITRS Combination Center at DGFI: a terrestrial reference frame realization 2003. *Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften (München), Reihe B: Angewandte Geodäsie (ISSN 0065-5317), Heft Nr. 313, p. 1 - 141 (2004), ISBN 3-7696-8593-8*, *313*, 1–141.

Argus, D., & Gordon, R. (1991). No-net-rotation model of current plate velocities incorporating plate motion model nuvel-1. *Geophysical Research Letters*, *18*. https://doi.org/10.1029/91GL01532

Arias, E. F., Charlot, P., Feissel, M., & Lestrade, J. .-. (1995). The extragalactic reference system of the International Earth Rotation Service, ICRS. *Astronomy & Astrophysics*, *303*, 604–608.

Artz, T., Springer, A., & Nothnagel, A. (2014). A complete VLBI delay model for deforming radio telescopes: the Effelsberg case. *J Geod*, *88*, 1145–1161. https://doi.org/10.1007/s00190-014-0749-1

Baars, J. (2007). *The paraboloidal reflector antenna in radio astronomy and communication*. Springer.

Bachmann, S., Thaller, D., Roggenbuck, O., Lösler, M., & Messerschmitt, L. (2016). Ivs contribution to itrf2014. *J Geod*, *90*, 631–654. https://doi.org/10.1007/s00190-016-0899-4

Bangert, J., Bell, S., Calabretta, M., Capitaine, N., Folkner, W., Hohenkerk, C., Li, J., Luzum, B., Malkin, Z., Percival, J., Ransom, S., & Wallace, P. (2017). *SOFA Time Scales and Calendar Tools*. IAU Standards of Fundamental Astronomy.

Bare, C., Clark, B. G., Kellermann, K. I., Cohen, M. H., & Jauncey, D. L. (1967). Interferometer experiment with independent local oscillators. *Science*, *157*(3785), 189–191.

Baver, K., Gipson, J., Carter, M. S., & Kingham, K. (2012). Assessment of the First Use of the Uniform Sky Strategy in Scheduling the Operational IVS-INT01 Sessions. In D. Behrend & K. Baver (Eds.), *IVS 2012 General Meeting Proceedings* (pp. 251–255). NASA/CP-2012-217504.

Behrend, D. (2013). Data Handling within the International VLBI Service. *Data Science Journal*, *12*, WDS81–WDS84. https://doi.org/10.2481/dsj.WDS-011

Beyer, W. (1985). *Laufzeitschätzung in digitalen Korrelationsinterferometern unter besonderer Betrachtung des MK-II-Systems des Max-Planck-Instituts für Radioastronomie* (Doctoral dissertation). Rheinische Friedrich-Wilhelms-Universität Bonn.

Böckmann, S., Artz, T., & Nothnagel, A. (2010). VLBI terrestrial reference frame contributions to ITRF2008. *J Geod*, *84*(3), 201–219. https://doi.org/10.1007/s00190-009-0357-7

Böhm, J., Böhm, S., Boisits, J., Girdiuk, A., Gruber, J., Hellerschmied, A., Krásná, H., Landskron, D., Madzak, M., Mayer, D., McCallum, J., McCallum, L., Schartner, M., & Teke, K. (2018). Vienna VLBI and Satellite Software (VieVS) for geodesy and astrometry [doi:10.1088/1538-3873/aaa22b]. *Publications of the Astronomical Society of the Pacific*, *130*(986), 044503.

Böhm, J. [J.], Salstein, D., Alizadeh, M. M., Wijaya, D. D., & Schuh, H. (2013). Geodetic and atmospheric background. In J. Böhm & H. Schuh (Eds.), *Atmospheric effects in space geodesy* (pp. 73–136). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-36932-2_3

Böhm, J. [J.], & Schuh, H. (2004). Vienna mapping functions in VLBI analyses. *GEOPHYS RES LETT*, *Vol. 31*.

Böhm, J. [J.], Werl, B., & Schuh, H. (2006). Troposphere mapping functions for GPS and very long baseline interferometry from European Centre for Medium-range Weather Forecasts operational analysis data. *Journal of Geophysical Research*, *111*, B02406. https://doi.org/10.1029/2005JB003629

Bolotin, S. (2000). VLBI data analysis center at main astronomical observatory of National Academy of Science of Ukraine. In N. Capitaine (Ed.), *Journées systèmes de référence spatio-temporels 2000* (pp. 102–106).

Boucher, C., & Altamimi, Z. (1990). *Evolution of the realizations of the Terrestrial Reference System done by the BIH and IERS (1984-1988)* (IERS Technical Note No. 4). International Earth Rotation and Reference Systems Service. Central Bureau of IERS - Observatoire de Paris.

Brockmann, E. (1997). *Combination of Solutions for Geodetic and Geodynamic Applications of the Global Positioning System (GPS)* (Dissertation No. 55). Schweizerische Geodätische Kommission. Geodätisch-geophysikalische Arbeiten. https://www.sgc.ethz.ch/sgc-volumes/sgk-55.pdf

Broten, N. W., Legg, T. H., Locke, J. L., McLeish, C. W., Richards, R. S., Chisholm, R. M., Gush, H. P., Yen, J. L., & Galt, J. A. (1967). Long base line interferometry: A new technique. *Science*, *156*(3782), 1592–1593. https://doi.org/10.1126/science.156.3782.1592

Brunner, F. K., & Gu, M. (1991). An improved model for the dual frequency ionospheric correction of GPS observations. *Manuscr. Geod.*, *16*(3), 205–214.

Burke, B. F., Graham-Smith, F., & Wilkinson, P. N. (2019). *An introduction to radio astronomy* (4th ed.). Cambridge University Press. https://doi.org/10.1017/9781316987506

Capitaine, N. (1990). The celestial pole coordinates. *Celestial Mech Dyn Astr*, *48*, 127–143. https://doi.org/10.1007/BF00049510

Capitaine, N., Guinot, B., & McCarthy, D. D. (2000). Definition of the Celestial Ephemeris Origin and of UT1 in the International Celestial Reference Frame. *AA*, *355*, 398–405.

Capitaine, N., Guinot, B., & Souchay, J. (1986). A non-rotating origin on the instantaneous equator: Definition, properties and use. *Celestial Mech Dyn Astr*, *39*, 283–307. https://doi.org/10.1007/BF01234311

Capitaine, N., & Wallace, P. T. (2006). High precision methods for locating the celestial intermediate pole and origin. *A&A*, *450*(2), 855–872. https://doi.org/10.1051/0004-6361:20054550

Cappallo, R. (2015). *Covariance analysis of the simultaneous fit of group delay and dtec in fourfit.* (Haystack Observatory Postprocessing System (HOPS) Memos). MIT. Haystack Observatory, Westford, Ma. https://www.haystack.mit.edu/haystack-observatory-postprocessing-system-hops/docs_hops_014_simul_ion_fit/

Cappallo, R. (2016). Delay and Phase Calibration in VGOS Post-Processing. In D. Behrend, K. D. Baver, & K. L. Armstrong (Eds.), *New horizons with vgos* (pp. 61–64).

Charlot, P. (1990). Radio-Source Structure in Astrometric and Geodetic Very Long Baseline Interferometry [doi:10.1086/115419]. *Astronomical Journal*, *99*, 1309. https://doi.org/10.1086/115419

Charlot, P., Jacobs, C. S., Gordon, D., Lambert, S., de Witt, A., Böhm, J., Fey, A. L., Heinkelmann, R., Skurikhina, E., Titov, O., Arias, E. F., Bolotin, S., Bourda, G., Ma, C., Malkin, Z., Nothnagel, A., Mayer, D., MacMillan, D. S., Nilsson, T., & Gaume, R. (2020). The third realization of the International Celestial Reference Frame by very long baseline interferometry. *A&A*, *644*, Article A159, A159. https://doi.org/10.1051/0004-6361/202038368

Clark, T. A. (1996). Low-cost gps time synchronization: The "totally accurate clock". In G. Beutler, W. G. Melbourne, G. W. Hein, & G. Seeber (Eds.), *Gps trends in precise terrestrial, airborne, and spaceborne applications* (pp. 325–327). Springer Berlin Heidelberg.

Clark, T. A., Corey, B. E., Davis, J. L., Elgered, G., Herring, T., Hinteregger, H. F., Knight, C., Levine, J., Lundqvist, G., Ma, C., Nesman, E., Phillips, R., Rogers, A., Ronnang, B., Ryan, J., Schupler, B., Shaffer, D., Shapiro, I., Vandenberg, N., ... Whitney, A. (1985). Precision Geodesy Using the Mark-III Very-Long-Baseline Interferometer System. *IEEE Transactions on Geoscience and Remote Sensing*, *GE-23*, 438–449. https://doi.org/10.1109/TGRS.1985.289433

Clark, T. A., & Thomsen, P. (1988). *Deformations in VLBI antennas* (NASA Technical Report NASA/TR 100696). NASA. NASA Greenbelt MD.

Cohen, M. H., & Shaffer, D. B. (1971). Positions of radio sources from long-baseline interferometry. *Astronomical Journal*, *76*, 76–91.

Collioud, A., & Charlot, P. (2009). The Bordeaux VLBI Image Database [https://ui.adsabs.harvard.edu/abs/2009evga.conf...19C/abstract]. In G. Bourda, P. Charlot, & A. Collioud (Eds.), *19th european vlbi for geodesy and astrometry working meeting* (pp. 19–22).

Corey, B. (2000). *Conversion from Mark IV to Mark III delays and rates - draft, january 18, 2000.* (MIT Haystack Observatory Report). MIT. Haystack Observatory, Westford, Ma.

Corey, B. (2006). *Note on antenna polarization and vlbi observables.* (IVS Memorandum 2007-011v01). MIT. Haystack Observatory, Westford, Ma. https://ivscc.gsfc.nasa.gov/publications/memos/ivs-2007-011v01.pdf

Cotton, W. D. (1993). Calibration and imaging of polarization sensitive Very Long Baseline Interferometer observations. *Astronomical Journal*, *106*(3), 1241–1248.

Davis, J. L., Herring, T. A., Shapiro, I. I., Rogers, A. E. E., & Elgered, G. (1985). Geodesy by radio interferometry - Effects of atmospheric modeling errors on estimates of baseline length. *Radio Science*, *20*(6), 1593–1607. https://doi.org/10.1029/RS020i006p01593

De Boor, C. (1978). *A practical guide to splines* (Vol. 27). Springer, New York.

Dehant, V., & Mathews, P. M. (2015). *Precession, Nutation and Wobble of the Earth*. Cambridge University Press.

Deller, A. T., Brisken, W. F., Phillips, C. J., Morgan, J., Alef, W., Cappallo, R., Middelberg, E., Romney, J., Rottmann, H., Tingay, S. J., & Wayth, R. (2011). DiFX-2: A More Flexible, Efficient, Robust, and Powerful Software Correlator [doi:10.1086/658907]. *Publications of the Astronomical Society of the Pacific*, *123*(901), 275. https://doi.org/10.1086/658907

Demets, C., Gordon, R., Argus, D., & Stein, S. (1990). Current plate motions. *Geophysical Journal International*, *101*, 425–478. https://doi.org/10.1111/j.1365-246X.1990.tb06579.x

de Witt, A., Charlot, P., Gordon, D., & Jacobs, C. S. (2022). Overview and Status of the International Celestial Reference Frame as Realized by VLBI. *Universe*, *8*(7). https://doi.org/10.3390/universe8070374

Dodson, A., Shardlow, P., Hubbard, L. C. M., Elgered, G., & Jarlemark, P. O. J. (1996). Wet tropospheric effects on precise relative GPS height determination. *J Geod*, *70*, 188–202. https://doi.org/10.1007/BF00873700

Elgered, G. (1982). Tropospheric wet path-delay measurements. *IEEE Trans. Antennas Prop.*, *30*(3), 502–505.

Elgered, G. (1993). Tropospheric radio path delay from ground-based microwave radiometry. In M. Janssen (Ed.), *Atmospheric remote sensing by microwave radiometry*. John Wiley & Sons, Inc., New York.

Eubanks, T. M. (1991). A Consensus Model for Relativistic Effects in Geodetic VLBI. In T. M. Eubanks (Ed.), *Proceedings of the U.S. Naval Observatory Workshop on Relativistic Models For Use in Space Geodesy* (pp. 60–82).

Fey, A. L., Gordon, D., Jacobs, C. S., Ma, C., Gaume, R. A., Arias, E. F., Bianco, G., Boboltz, D. A., Böckmann, S., Bolotin, S., Charlot, P., Collioud, A., Engelhardt, G., Gipson, J., Gontier, A.-M., Heinkelmann, R., Kurdubov, S., Lambert, S., Lytvyn, S., … Zharov, V. (2015). The Second Realization of the International Celestial Reference Frame by Very Long Baseline Interferometry. *The Astronomical Journal*, *150(2)*(58). https://doi.org/10.1088/0004-6256/150/2/58

Fricke, W., Schwan, H., Lederle, T., Bastian, U., Bien, R., Burkhardt, G., Du Mont, B., Hering, R., Jährling, R., Jahreiß, H., Röser, S., Schwerdtfeger, H. .-., & Walter, H. G. (1988). Fifth Fundamental Catalogue (FK5). Part 1. The Basic Fundamental Stars. *Veroeffentlichungen des Astronomischen Rechen-Instituts Heidelberg*, *32*, 1–106.

Gipson, J. (2016). *Sked - VLBI scheduling software*. NASA Goddard Spaceflight Center. https://lupus.gsfc.nasa.gov/files_user_manuals/sked/SkedManual_v2016Dec09.pdf

Gipson, J., MacMillan, D., & Petrov, L. (2008). Improved Estimation in VLBI through Better Modeling and Analysis. In A. Finkelstein & D. Behrend (Eds.), *IVS General Meeting Proceedings* (pp. 157–162).

Grafarend, E. W., & Sansó, F. (1984). *Optimization and design of geodetic networks*. Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-642-70659-2

Guinot, B. (1979). Basic Problems in the Kinematics of the the Rotation of the Earth. In D. McCarthy & J. Pilkington (Eds.), *Time and the earth's rotation* (pp. 7–18).

Halsig, S., Artz, T., Iddink, A., & Nothnagel, A. (2016). Using an atmospheric turbulence model for the stochastic model of geodetic vlbi data analysis. *Earth, Planets and Space*, *68:106*. https://doi.org/10.1186/s40623-016-0482-5

Hamaker, J., & Bregman, J. (1996). Understanding radio polarimetry. iii. interpreting the iau/ieee definitions of the stokes parameters. *Astronomy and Astrophysics Supplement Series*, *117*, 161–165. https://doi.org/10.1051/aas:1996147

Hamaker, J., Bregman, J., & Sault, R. (1996). Understanding radio polarimetry. i. mathematical foundations. *Astronomy and Astrophysics Supplement Series*, *117*, 137–147. https://doi.org/10.1051/aas:1996146

Hase, H., & Petrov, L. (1999). The First Campaign of Observations with VLBI-Module of TIGO. In W. Schlüter & H. Hase (Eds.), *Proceedings of the 13th Working Meeting on European VLBI for Geodesy and Astrometry, Viechtach, Germany* (pp. 19–24). BKG.

Hellmers, H., Modiri, S., Bachmann, S., Thaller, D., Bloßfeld, M., Seitz, M., & Gipson, J. (2022). Combined ivs contribution to the itrf2020. Springer Berlin Heidelberg. https://doi.org/10.1007/1345_2022_170

Helmert, F. (1872). *Die ausgleichungsrechnung nach der methode der kleinsten quadrate*. Teubner, Leipzig.

Herring, T. A. (1983). *Precision and accuracy of intercontinental distance determinations using radio interferometry* (Dissertation AFGL-TR-84-0182). Hanscom AFB, Massachusetts 01731, USA. MIT, 77 Massachusetts Av., Cambridge MA 02139, USA, Hanscom AFB, Massachusetts 01731, USA.

Herring, T. A. (1992). Modeling Atmospheric Delays in the Analysis of Space Geodetic Data. In J. De Munck & T. Spoelstra (Eds.), *Refraction of transatmospheric signals in geodesy* (pp. 157–164). Netherlands Geodetic Commission Publications on Geodesy.

Herring, T. A., Davis, J. L., & Shapiro, I. I. (1990). Geodesy by radio interferometry: The application of Kalman filtering to the analysis of very long baseline interferometry data. *Journal of Geophysical Research*, *95*(B8), 12561–12581. https://doi.org/10.1029/JB095iB08p12561

Herring, T. A., Shapiro, I. I., Clark, T. A., Ma, C., Ryan, J. W., Schupler, B. R., Knight, C. A., Lundqvist, G., Shaffer, D. B., Vandenberg, N. R., Corey, B. E., Hinteregger, H. F., Rogers, A. E. E., Webber, J. C., Whitney, A. R., Elgered, G., Rönnäng, B., & Davis, J. L. (1986). Geodesy by radio interferometry: Evidence for contemporary plate motion. *Journal of Geophysical Research*, *91*, 8341–8347. https://api.semanticscholar.org/CorpusID:128905578

Himwich, E., Bertarini, A., Corey, B., Baver, K., Gordon, D., & La Porta, L. (2017). Impact of Station Clocks on UT1-TAI Estimates. In R. Haas & G. Elgered (Eds.), *Proceedings of the 23rd Working Meeting on European VLBI for Geodesy and Astrometry* (pp. 172–176). Chalmers University of Technology.

Hinteregger, H. F., Shapiro, I. I., Robertson, D. S., Knight, C. A., Ergas, R. A., Whitney, A. R., Rogers, A. E. E., Moran, J. M., Clark, T. A., & Burke, B. F. (1972). Precision geodesy via radio interferometry. *Science*, *178*, 396–398. https://doi.org/10.1126/science.178.4059.396

Hobiger, T., & Kondo, T. (2005). An FX software correlator based on Matlab. *Proceedings of the 17th Working Meeting on European VLBI for Geodesy and Astrometry*, 34–38.

Hobiger, T., Koyama, Y., Boehm, J., Kondo, T., & Ichikawa, R. (2009). The effect of neglecting VLBI reference station clock offsets on UT1 estimates. *Advances in Space Research*, *43*(6), 910–916. https://doi.org/https://doi.org/10.1016/j.asr.2008.11.005

Holst, C., Nothnagel, A., Blome, M., Becker, P., Eichborn, M., & Kuhlmann, H. (2015). Improved area-based deformation analysis of a radio telescope's main reflector based on terrestrial laser scanning. *J Applied Geodesy*, *9*(1), 1–13. https://doi.org/10.1515/jag-2014-0018

Holst, C., Nothnagel, A., Haas, R., & Kuhlmann, H. (2019). Investigating the gravitational stability of a radio telescope's reference point using a terrestrial laser scanner: Case study at the Onsala Space Observatory 20-m radio telescope. *ISPRS Journal of Photogrammetry and Remote Sensing*, *149*, 67–76. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2019.01.010

Jacobs, C., Heflin, M., Lanyi, G., Sovers, O., & Steppe, J. (2010). Rotational Alignment Altered by Source Position Correlations. *Sixth International VLBI Service for Geodesy and Astronomy. Proceedings from the 2010 General Meeting, "VLBI2010: From Vision to Reality". Held 7-13 February, 2010 in Hobart, Tasmania, Australia. Edited by D. Behrend and K.D. Baver. NASA/CP 2010-215864., p.305-309*, 305–309.

Johansson, L., Stodne, F., & Wolf, S. (1996). *The PISA project - variations in the height of the foundation of the 20 meter radio telescope* (Technical Report No. 178). Chalmers Technical University, Gothenburg, Sweden. Onsala Space Observatory, Sweden, Chalmers Technical University, Gothenburg, Sweden.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, *82(1)*, 35–45. https://doi.org/10.1115/1.3662552

Karbon, M., Heinkelmann, R., Mora-Diaz, J., Xu, M., Nilsson, T., & Schuh, H. (2017). The extension of the parametrization of the radio source coordinates in geodetic VLBI and its impact on the time series analysis. *Journal of Geodesy*, *91*, 755–765. https://doi.org/10.1007/s00190-016-0954-1

Koch, K.-R. (1999). *Parameter estimation and hypothesis testing in linear models*. Springer Verlag, Berlin, Paris, Heidelberg, New York.

Kondo, T., & Takefuji, K. (2016). An algorithm of wideband bandwidth synthesis for geodetic VLBI. *Radio Science*, *51*(10), 1686–1702. https://doi.org/https://doi.org/10.1002/2016RS006070

Kondo, T., & Takefuji, K. (2019). *Current status of tec estimation in the wideband bandwidth synthesis software.* (NICT-TDC News No. 38). NICT. 4-2-1 Nukui-kita, Koganei, Tokyo 184-8795, Japan. https://www2.nict.go.jp/sts/stmg/ivstdc/news_38/pctdc_news38.pdf

Krásná, H., Nickola, M., & Böhm, J. (2014). Axis Offset Estimation of VLBI Telescopes. In D. Behrend, K. Baver, & K. Armstrong (Eds.), *IVS 2014 General Meeting Proceedings* (pp. 339–343). Science Press (Beijing) ISBN 978-7-03-042974-2.

Kurdubov, S., & Skurikhina, E. (2010). Antenna Axis Offset Estimation from VLBI. In D. Behrend & K. Baver (Eds.), *IVS 2010 General Meeting Proceedings* (pp. 247–250). NASA/CP-2010-215864.

Lambeck, K. (1980). *The Earth's Variable Rotation*. Cambridge University Press.

Landskron, D., & Böhm, J. [J.]. (2018). VMF3/GPT3: Refined discrete and empirical troposphere mapping functions. *J Geod*, *92*(4), 349–360. https://doi.org/DOI10.1007/s00190-017-1066-2

Lieske, J. H. (1979). Precession matrix based on IAU (1976) system of astronomical constants. *A & A*, *73*(3), 282–284.

Lösler, M., Haas, R., & Eschelbach, C. (2013). Automated and continual determination of radio telescope reference points with sub-mm accuracy: Results from a campaign at the Onsala Space Observatory. *J Geod*, *87*(8), 791–804. https://doi.org/10.1007/s00190-013-0647-y

Lösler, M., Haas, R., & Eschelbach, C. (2016). Terrestrial monitoring of a radio telescope reference point using comprehensive uncertainty budgeting – in- vestigations during cont14 at the onsala space observatory. *JoG*, *90*(5), 467–486. https://doi.org/10.1007/s00190-016-0887-8

Ma, C., & Feissel, M. (1997). *Definition and realization of the International Celestial Reference System by VLBI astrometry of extragalactic objects.* (IERS Technical Note No. 23) [ISSN: 1019-4568]. International Earth Rotation and Reference Systems Service. Central Bureau of IERS - Observatoire de Paris.

Ma, C., Ryan, J., & Caprette, D. (1989). *Crustal dynamics project data analysis - 1988 vlbi geodetic results 1979-87* (Technical Report). NASA Crustal Dynamics Project. NASA Goddard Space Flight Center, Greenbelt, Maryland.

MacMillan, D. S. (1995). Atmospheric gradients from very long baseline interferometry observations. *Geophysical Research Letters*, *22*(9), 1041–1044. https://doi.org/10.1029/95gl00887

Marini, J., & Murray, W. (1973). Correction of laser range tracking data for atmospheric refraction at elevation angles above 10 degrees. *NASA Technical Report X-591-73-351, Goddard Space Flight Center, Greenbelt MD*.

Martí-Vidal, I., Mus, A., Janssen, M., de Vicente, P., & González, J. (2021). Polarization calibration techniques for the new-generation vlbi. *Astronomy & Astrophysics*, *646*, A52. https://doi.org/10.1051/0004-6361/202039527

Martí-Vidal, I., Roy, A., Conway, J., & Zensus, A. J. (2016). Calibration of mixed-polarization interferometric observations - tools for the reduction of interferometric data from elements with linear and circular polarization receivers. *A&A*, *587*, A143. https://doi.org/10.1051/0004-6361/201526063

Mathews, P. M., Herring, T. A., & Buffett, B. A. (2002). Modeling of nutation and precession: New nutation series for nonrigid earth and insights into the earth's interior. *Journal of Geophysical Research: Solid Earth*, *107*(B4), ETG 3-1-ETG 3–26. https://doi.org/https://doi.org/10.1029/2001JB000390

Matveenko, L. I. (2007). Early vlbi in the ussr. *Astronomische Nachrichten*, *328*, 411–419. https://doi.org/10.1002/asna.200710763

Matveenko, L. I., Kardashev, N. S., & Sholomitskii, G. B. (1965). Large base-line radio interferometers. *Soviet Radiophysics*, *8*(4), 461–463. https://doi.org/10.1007/bf01038318

McCallum, L., Chuan, L., Krásná, H., McCallum, J., Böhm, J., McCarthy, T., Gruber, J., Schartner, M., Quick, J., & Rogers, A. (2022). The australian mixed-mode observing program. *J Geod*, *96*, 67. https://doi.org/10.1007/s00190-022-01657-2

McCarthy, D. (1992). *IERS Standards 1992* (IERS Technical Note 13). International Earth Rotation and Reference Systems Service. Central Bureau of IERS - Observatoire de Paris.

McCarthy, D. (1996). *IERS Conventions 1992* (IERS Technical Note 21). International Earth Rotation and Reference Systems Service. Central Bureau of IERS - Observatoire de Paris.

McCarthy, D., & Luzum, B. (2004). *IERS Conventions 2003* (IERS Technical Note No. 32) [ISSN: 1019-4568]. International Earth Rotation and Reference Systems Service. Verlag des Bundesamtes für Kartographie und Geodäsie, Frankfurt am Main.

Michelson, A. A. (1881). The relative motion of the Earth and of the luminiferous ether. *American Journal of Science*, *22*(128), 120–129. https://doi.org/10.2475/ajs.s3-22.128.120

Michelson, A. A. (1890). I. on the application of interference methods to astronomical measurements. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *30*(182), 1–21. https://doi.org/10.1080/14786449008619983

Minster, B., & Jordan, T. (1978). Present day plate motions. *J. Geophys. Res.*, *83*. https://doi.org/10.1029/JB083iB11p05331

Moran, J. M., Crowther, P. P., Burke, B. F., Barrett, A. H., Rogers, A. E. E., Ball, J. A., Carter, J. C., & Bare, C. C. (1967). Spectral line interferometry with independent time standards at stations separated by 845 Kilometers. *Science*, *157*(3789), 676–677. https://doi.org/10.1126/science.157.3789.676

Mueller, I. I. (1969). *Spherical and Practical Astronomy as Applied to Geodesy*. Frederic Ungar Publishing Co., New York.

Niell, A. E. (1996). Global mapping functions for the atmosphere delay at radio wavelengths. *Journal of Geophysical Research*, *101*(B02), 3227–3246. https://doi.org/10.1029/95JB03048

Niell, A. E. (2000). Improved atmospheric mapping functions for VLBI and GPS. *Earth, Planets and Space*, *52*(10), 699–702. https://doi.org/10.1186/bf03352267

Niell, A. E., Barrett, J., Burns, A., Cappallo, R., Corey, B., Derome, M., Eckert, C., Elosegui, P., McWhirter, R., Poirier, M., Rajagopalan, G., Rogers, A., Ruszczyk, C., SooHoo, J., Titus, M., Whitney, A., Behrend, D., Bolotin, S., Gipson, J., … Petrachenko, B. (2018). Demonstration of a broadband very long baseline interferometer system: A new instrument for high-precision space geodesy [doi:10.1029/2018RS006617]. *Radio Science*, *53*(10), 1269–1291.

Niell, A. E., Barrett, J. P., Cappallo, R. J., Corey, B. E., Elosegui, P., Mondal, D., Rajagopalan, G., Ruszczyk, C. A., & Titus, M. A. (2021). VLBI measurement of the vector baseline between geodetic antennas at Kokee Park Geophysical Observatory, Hawaii. *J Geod*, *95*, 65. https://doi.org/10.1007/s00190-008-0284-z

Niell, A. E., Whitney, A., Petrachenko, B., Schlüter, W., Vandenberg, N., Hase, H., Koyama, Y., Ma, C., Schuh, H., & Tuccari, G. (2006). VLBI2010: Current and future requirements for geodetic VLBI systems [NASA/TP-2006-214136]. In D. Behrend & K. Baver (Eds.), *IVS 2005 Annual Report* (pp. 13–40).

Niemeier, W. (2002). *Ausgleichungsrechnung*. de Gruyter.

Nilsson, T., Mora-Diaz, J. A., Raposo-Pulido, V., Heinkelmann, R., Karbon, M., Liu, L., Lu, C., Xu, B. S. M., & Schuh, H. (2017). Antenna axis offsets and their impact on vlbi derived reference frames. In T. van Dam (Ed.), *Refag 2014* (pp. 53–58). Springer International Publishing.

Nilsson, T., Soja, B., Karbon, M., Heinkelmann, R., & Schuh, H. (2015). Application of Kalman filtering in VLBI data analysis. *Earth Planet Sp*, *67*(1). https://doi.org/10.1186/s40623-015-0307-y

Nothnagel, A. (2002). Summary of the Third IVS Analysis Workshop. In N. R. Vandenberg & K. D. Baver (Eds.), *International VLBI Service for Geodesy and Astrometry: General Meeting Proceedings, NASA/CP-2002-210002* (pp. 394–397).

Nothnagel, A. (2009). Conventions on thermal expansion modelling of radio telescopes for geodetic and astrometric VLBI. *J Geod*, *83*(8), 782–792. https://doi.org/10.1007/s00190-008-0284-z

Nothnagel, A., Artz, T., Behrend, D., & Malkin, Z. (2017). International VLBI Service for Geodesy and Astrometry - delivering high-quality products and embarking on observations of the next generation. *J Geod*, *91*(7), 711–721. https://doi.org/10.1007/s00190-016-0950-5

Nothnagel, A., & Krásná, H. (2021). Reference epochs in VLBI estimations of clock parameters. *Proceedings of the 25th Working Meeting on European VLBI for Geodesy and Astrometry*.

Nothnagel, A., Nilsson, T., & Schuh, H. (2018). Very Long Baseline Interferometry: Dependencies on Frequency Stability [doi:10.1007/s11214-018-0498-1]. *Space Science Reviews*, *214*(3), Article 66, 66. https://doi.org/10.1007/s11214-018-0498-1

Owens, J. (1967). Optical refractive index of air: Dependence on pressure, temperature and composition. *Applied Optics*, *Vol. 6, No. 1*, 51–59.

Petit, G., & Luzum, B. (2010). *IERS Conventions 2010* (IERS Technical Note No. 36) [ISSN: 1019-4568]. International Earth Rotation and Reference Systems Service. Verlag des Bundesamtes für Kartographie und Geodäsie, Frankfurt am Main.

Petrachenko, B., Bujold, M., Cannon, W. H., Carlson, B. R., Dewdney, P. E., Feil, G. H., Newby, P., Novikov, A., Popelar, J., & Wietfeldt, R. D. (2000). The s2 vlbi system: Das, rt/pt and correlator. In F. Takahashi (Ed.), *Very Long Baseline Interferometry (VLBI) International VLBI Service for Geodesy and Astrometry 2000 General Meeting Proceedings* (pp. 378–382).

Petrachenko, B., Niell, A., Behrend, D., Corey, B., Böhm, J., Charlot, P., Collioud, A., Gipson, J., Haas, R., Hobiger, T., Koyama, Y., MacMillan, D., Malkin, Z., Nilsson, T., Pany, A., Tuccari, G., Whitney, A., & Wresnik, J. (2009). *Design Aspects of the VLBI2010 System. Progress Report of the VLBI2010 Committee* (NASA Technical Memorandum NASA/TM-2009-214180). International VLBI Service for Geodesy and Astrometry. NASA Greenbelt MD.

Petrov, L. (1999). Steps towards phase delay VLBI. In W. Schlüter & H. Hase (Eds.), *Proceedings of the 13th Working Meeting on European VLBI for Geodesy and Astrometry, Viechtach, Germany* (pp. 144–151). BKG, Frankfurt a. M.

Petrov, L. (2001). *Memo: Ionophere contribution in VLBI* (Technical Memo) [http:// astrogeo.org/petrov/papers/iono.ps.gz]. AstroGeo.

Petrov, L., Gordon, D., Gipson, J., MacMillan, D., Ma, C., Fomalont, E., Walker, R. C., & Carabajal, C. (2009). Precise geodesy with the very long baseline array. *J Geod*, *83*(9), 859–876. https://doi.org/10.1007/s00190-009-0304-7

Petrov, L., Kovalev, Y. Y., Fomalont, E. B., & Gordon, D. (2011). The Very Long Baseline Array Galactic Plane Survey—VGaPS [doi:10.1088/0004-6256/142/2/35]. *Astronomical Journal*, *142*(2), Article 35, 35. https://doi.org/10.1088/0004-6256/142/2/35

Plag, H. P., & Pearlman, M. (2009). *Global geodetic observing system: Meeting the requirements of a global society on a changing planet in 2020*. Springer Verlag. https://doi.org/10.1007/978-3-642-02687-4_3

Pogrebenko, S. (1993a). *Proposal for a Multiband & Multitone Phase Calibration Signal Extractor for the EVN Upgrade Project* (EVN DOC 2). EVN.

Pogrebenko, S. (1993b). *Analysis of Phase Cal Extraction Algorithms and a Proposal for a Parallel Multitone Phase Cal Extraction Software Development* (EVN DOC 2). EVN.

Robertson, D. S. (1991). Geophysical applications of very long baseline interferometry. *Rev Mod Phys*, *63:4*, 899–918.

Rogers, A. E. E. (1970). Very long baseline interferometry with large effective bandwidth for phase-delay measurements. *Radio Science*, *5*(10), 1239–1247. https://doi.org/10.1029/RS005i010p01239

Rogers, A. E. E. (1980). Phase and group delay calibration of a very long baseline interferometer by east coast VLBI group. *Radio Interferometry Techniques for Geodesy; NASA Conference Publication 2115*, 255–261.

Romero-Wolf, A., Jacobs, C. S., & Ratcliff, J. T. (2012). Effects of Tropospheric Spatio-temporal Correlated Noise on the Analysis of Space Geodetic Data. *Seventh General Meeting (GM2012) of the international VLBI Service for Geodesy and Astrometry (IVS*, 231–235.

Saastamoinen, J. (1972). Atmospheric correction for the troposphere and stratosphere in radio ranging of satellites. *In: The Use of Artificial Satellites for Geodesy, Geophysic. Monogr. Ser., Vol. 15, S. W. Henriksen et al. (Eds), American Geophysical Union, Washington DC*.

Saastamoinen, J. (1973a). Contributions to the theory of atmospheric refraction. part i - astronomical refraction. *Bulletin Géodésique*, *105*(1), 279–298. https://doi.org/10.1007/BF02521844

Saastamoinen, J. (1973b). Contributions to the theory of atmospheric refraction. part ii - refraction ccorrection in satellite geodesy. *Bulletin Géodésique*, *105*(1), 279–298. https://doi.org/10.1007/BF02521844

Sarti, P., Abbondanza, C., & Vittuari, L. (2009). Gravity-dependent signal path variation in a large VLBI telescope modelled with a combination of surveying methods. *J Geod*, *83*(11), 1115–1126. https://doi.org/10.1007/s00190-009-0331-4

Schlüter, W., & Behrend, D. (2007). The International VLBI Service for Geodesy and Astrometry (IVS): Current capabilities and future prospects. *J Geod*, *81*(6-8), 379–387. https://doi.org/10.1007/s00190-006-0131-z

Schuh, H., & Behrend, D. (2012). VLBI: A fascinating technique for geodesy and astrometry. *Journal of Geodynamics*, 68–80. https://doi.org/10.1016/j.jog.2012.07.007

Shapiro, I. I., Robertson, D. S., Knight, C. A., Counselman, C. C., Rogers, A. E. E., Hinteregger, H. F., Lippincott, S., Whitney, A. R., Clark, T. A., Niell, A. E., & Spitzmesser, D. J. (1974). Transcontinental baseline and the rotation of the earth measured by radio interferometry. *Science*, *186*, 920–922.

Shuygina, N., Ivanov, D., Ipatov, A., Gayazov, I., Marshalov, D., Melnikov, A., Kurdubov, S., Vasilyev, M., Ilin, G., Skurikhina, E., Surkis, I., Mardyshkin, V., Mikhailov, A., Salnikov, A., Vytnov, A., Rakhimov, I., Dyakov, A., & Olifirov, V. (2019). Russian vlbi network “quasar”: Current status and outlook [From Space Geodesy to Astro-Geodynamics]. *Geodesy and Geodynamics*, *10*(2), 150–156. https://doi.org/https://doi.org/10.1016/j.geog.2018.09.008

Smirnov, O. M. (2011). Revisiting the radio interferometer measurement equation - i. a full-sky jones formalism. *A&A*, *527*, A106. https://doi.org/10.1051/0004-6361/201016082

Smith, E. K., & Weintraub, S. (1953). The constants in the equation for the atmospheric refractive index at radio frequencies. *Proc. IRE*, *41*, 1035–1037.

Soja, B., Heinkelmann, R., & Schuh, H. (2014). Probing the solar corona with very long baseline interferometry. *Nature Communications*, *5*(4166). https://doi.org/10.1038/ncomms5166

Soja, B., Nilsson, T., Karbon, M., Zus, F., Dick, G., Deng, Z., Wickert, J., Heinkelmann, R., & Schuh, H. (2015). Tropospheric delay determination by Kalman filtering VLBI data. *Earth Planet Space*, *67*(1). https://doi.org/10.1186/s40623-015-0293-0

Sovers, O. J., Fanselow, J. L., & Jacobs, C. S. (1998). Astrometry and geodesy with radio interferometry: experiments, models, results. *Rev Mod Phys*, *70*, 1393–1454. https://doi.org/10.1103/RevModPhys.70.1393

Steufmehl, H. J. (1994). *Optimierung von Beobachtungsplänen in der Langbasisinterferometrie (VLBI)* (Doctoral dissertation) [Deutsche Geodätische Kommission Bayer. Akad. Wiss. München, Reihe C, Vol. 406]. Rheinische Friedrich-Wilhelms-Universität Bonn.

Stokes, G. G. (1852). On the Composition and Resolution of Streams of Polarized Light from different Sources. *Proceedings of the Cambridge Philosophical Society : Mathematical and physical sciences, vol. 9.*, *9*, 399.

Sun, J., Böhm, J., Nilsson, T., Krásná, H., Böhm, S., & Schuh, H. (2014). New VLBI2010 scheduling strategies and implications on the terrestrial reference frames. *J Geod*, *88*(5), 449–461. https://doi.org/10.1007/s00190-014-0697-9

Takahashi, F., Kondo, T., Takahasi, Y., & Koyama, Y. (2000). *Wave summit course: Very long baseline interferometer*. Ohmsha Ltd., Tokyo, Japan.

Taylor, H., & Golomb, S. (1985). *Rulers, Part i* (CSI Technical Report No. 85-05-01). Communication Science Institute, Univ. Southern California, Los Angeles, CA.

Thayer, G. D. (1974). An improved equation for the radio refractive index of air. *Radio Science*, *Vol. 9, No. 10*, 803–807.

Thompson, A., Moran, J., & Swenson Jr., G. (2007). *Interferometry and synthesis in radio astronomy, 2nd edition* [ISBN: 978-0-471-25492-8]. John Wiley; Sons Ltd. https://doi.org/10.1002/9783527617845

Thompson, A., Moran, J., & Swenson Jr., G. (2017). *Interferometry and Synthesis in Radio Astronomy, 3rd Edition* [doi:10.1007/978-3-319-44431-4]. Springer. https://doi.org/10.1007/978-3-319-44431-4

Titov, O. A. (2000). Estimation of the subdiurnal UT1-UTC variations by the least squares collocation method. *Astronomical & Astrophysical Transactions*, *18*(6), 779–792. https://doi.org/10.1080/10556790008208172

Uunila, M., Haas, R., Kareinen, N., & Lindfors, T. (2012). Automated Analysis of dUT1 from IVS Intensive Sessions with VieVS. In D. Behrend &. K. Baver (Eds.), *IVS 2012 General Meeting Proceedings, Madrid, Spain, March 4 - 9 2012* (pp. 281–285). NASA/CP-2012-217504.

Van Vleck, J. H., & Middleton, D. (1966). The spectrum of clipped noise. *Proceedings of the IEEE*, *54*(1), 2–19. https://doi.org/10.1109/PROC.1966.4567

Varenius, E., Haas, R., & Nilsson, T. S. (2021). Short-baseline interferometry local-tie experiments at the Onsala Space Observatory. *J Geod*, *95*, 54. https://doi.org/10.1007/s00190-021-01509-5

Vennebusch, M., Böckmann, S., & Nothnagel, A. (2007). The contribution of Very Long Baseline Interferometry to ITRF2005. *J Geod*, *81*, 553–564. https://doi.org/10.1007/s00190-006-0117-x

Verhoeven, G. (2017). he reflection of two fields – Electromagnetic radiation and its role in (aerial) imaging. *AARGnews*, *55*, 13–18. https://doi.org/10.5281/zenodo.3534245

Whitney, A. R. (1974). *Precision geodesy and astrometry via very-long-baseline interferometry* (Dissertation). MIT. MIT, 77 Massachusetts Av., Cambridge MA 02139, USA, Dissertation, MIT.

Whitney, A. R. (2000). How Do VLBI Correlators Work? In. N. Vandenberg &. K. Baver (Eds.), *IVS 2000 General Meeting Proceedings, Bad Kötzting, Germany, February 21-24 2000* (pp. 187–205). NASA/CP-2000-209893.

Yang, J., Kildal, P., & Pantaleev, M. (2013). Overview of developments of the Eleven Feeds. In *7th European Conference on Antennas and Propagation, EuCAP, 8-12 April 2013* (pp. 553–557). Gothenburg, Sweden.

Zernecke, R. (1999). Seasonal variations in height demonstrated at the radio telescope reference point. In W. Schlüter & H. Hase (Eds.), *Proceedings of the 13th Working Meeting on European VLBI for Geodesy and Astrometry, Viechtach, Germany* (pp. 15–18). BKG.

List of Abbreviations and Variables

AO Antenna Axis Offset

AP Accumulation Period

BCO Baseline clock offset

BWS Bandwidth synthesis

CDMS Cable Delay Measurement System

CIO Celestial Intermediate Origin

CIP Celestial Intermediate Pole

CPO Celestial Pole Offset

CRF Celestial Reference Frame

DBBC Digital BaseBand Converter

GCRS Geocentric Celestial Reference System

GCRS Geocentric Celestial Reference System

GGOS Global Geodetic Observing System

HOPS Haystack Observatory Postprocessing System

IAG International Association of Geodesy

IAU International Astronomical Union

ICRF International Celestial Reference Frame

ICRS International Celestial Reference System

IF Intermediate Frequency

ITRF International Terrestrial Reference Frame

ITRS International Terrestrial Reference System

ITU International Telecommunications Union

IVS International VLBI Service for Geodesy and Astrometry

Jy Jansky (Flux density)

LCP Left hand circular polarization

LNA Low noise amplifier

LSB Lower Side Band

MBD Multiband Delay

NRO Non-rotating Origin

OPC Observing Program Committee

PO Peculiar Offset

QRFH Quad Ridge Flared Horn

RCP Right hand circular polarization

RDBE Reconfigurable Open Architecture Computing Hardware (ROACH) Digital Backend

SBD Single Band Delay

SNR Signal to Noise Ratio

TEC(U) Total Electron Content (Units)

TIO Terrestrial Intermediate Origin

TRF Terrestrial Reference Frame

UDC Up/Down converter

USB Upper Side Band

VGOS VLBI Global Observing System

VLBI Very Long Baseline Interferometry

ZWD Zenith Wet Delay

$(1),(2)$	Telescopes # 1 and # 2
$\mathbf{c}$	Speed of light
ν_0	Observing Frequency in Hz
ω	Angular Frequency in radian
τ_g	A priori total group delay from geometric model
τ_{gr}	Observed total group delay
$\Box\tau_{gr}$	Group delay ambiguity spacing
$\Box\tau_{ph}$	Phase delay ambiguity spacing

Glossary

- **Band**: More correctly frequency band, denotes a block of frequencies in the electromagnetic spectrum. A few common examples are L band (according to ITU 1 - 2 GHz), S band (2 - 4 GHz), C band (4 - 8 GHz), X band (8 - 12 GHz), Ku band (12 - 18 GHz), K band (18 - 27 GHz), or Ku band (27 - 40 GHz).
  In VGOS terminology, there are bands A, B, C, and D, which denote special regions of the spectrum.
- **Bandpass**: The form of the power pattern with respect to frequency
- **Baseband**: Frequency band starting at 0 Hz (DC) up to some finite frequency mostly in the MHz regime.
- **Channel**: See sub-band.
- **Global solution**: A least squares solution of all or most available geodetic and astrometric VLBI observing sessions is called a global solution. Sometimes this can be restricted to a certain type of sessions.
- **Level-0 Data**: The raw digitized noise gathered at the radio telescopes.
- **Level-0 Data Analysis**: The correlation process. The output are the fringe visibilities.
- **Level-1 Data**: Fringe visibilities/phasors.
- **Level 1 Data Analysis**: Produces the observables phase and group delays as well as their time derivatives. Also included are all necessary analysis steps at this stage, such as polarization combination and fringe fitting of the visibilities.
- **Level-2 Data**: The observables phase and group delays as well as their time derivatives.
- **Level-2 Data Analysis**: The analysis steps working with phase and group delays and their rates for the computation of geodetic and astrometric parameters. This also includes work on source imaging and source structure effects.
- **Level-3 Data**: Geodetic and astrometric parameters.
- **Level-3 Data Analysis**: The combination of several Level-3 data sets, e.g., from different analysis centers.
- **Local oscillator**: The aggregate components to generate one or more frequencies for mixing down the frequency band received down to baseband.
- **Sub-band**: Part of a frequency band which is processed in a separate channel of the signal chain.

- **Phasor**: A phasor is the representation of amplitude and phase as a complex number.
- **vgosDB**: The name of a suite of files, also called database, in the standard geodetic VLBI data storage format based on NetCDF data storage format[33].

[33] https://docs.unidata.ucar.edu/netcdf-c/current/

Index

Acknowledgements

Many individuals helped me with their advice and recommendations, and I am grateful for their expertise, time and patience. Almost all of the section on correlation was written by Thomas Hobiger of Chalmers University of Technology, Gothenburg, Sweden. With Sigrid Böhm and Jan Vondrak I discussed nutation modelling. Arthur Niell was involved in several sections. He explained to me details on HOPS *fourfit* and on system calibrations, and checked the VGOS section suggesting improvements for a better understanding of the background. We also had long discussions on (Level 2) phase solutions. Brian Corey provided insights into HOPS *fourfit*, the pseudo-Stokes I (Ixy) approach, and many, many other topics. Roger Capallo helped me to figure out what is going on in *fourfit* with respect to the phase observable. Sergei Bolotin explained to me how phases are handled in *nuSolve*. Ivan Martí-Vidal provided valuable comments to my drafts and taught me a lot on polarization combinations for VGOS. Many thanks to Frédéric Jaron once again for meticulously checking text passages especially on polarization and VGOS fringe fitting. Thanks as well to Geert Verhoeven, Ludwig Boltzmann Institute for Archaeological Prospection and Virtual Archaeology, Wien, Austria, for preparing and providing his graphs as vector plots. I also thank Hana Krásná, Manuela Seitz, and Lisa Kern for the discussions on datum definition as well as Helene Wolf for frequently giving advice on LaTeX issues.