Chapter 7

Band pass and Phase Calibration

Robert Lucas

IRAM, 300 rue de la Piscine, F-38406 Saint Martin d'Hères

7.1 Definitions and formalism

As has been seen in the previous lectures, each interferometer baseline provides a measurement of the source visibility at a given point in the u, v plane of spatial frequencies; the source brightness distribution can then be reconstructed by an appropriate Fourier Transform.

In reality things are not so simple. Interferometers are designed with a lot of care; however many electronic components will have variable gains both in amplitude and in phase; these variations will affect the results and have to be taken out. It is generally sometimes more efficient to have a slightly varying instrument response, and a more sensitive instrument, than a very stable one with less sensitivity, provided the varying terms in the response are slow and may be easily calibrated out. At millimeter wavelengths the atmospheric absorption and path length fluctuations will dominate the instrument imperfections in most cases.

For a given observation, if we interpret the correlator response (amplitude and phase) as the source visibility, ignoring any imperfections, we have an observed (apparent) visibility $\widetilde{V}_{ij}(t)$, where i, j are antenna numbers, ν the frequency and t is time. If the true source visibility is $V_{ij}(t)$, we may define :

$$\widetilde{V}_{ij}(t) = \mathcal{G}_{ij}(t) V_{ij}(t) + \epsilon_{ij}(t) + \eta_{ij}(t) \tag{7.1}$$

where the $\mathcal{G}_{ij}(\nu, t)$ are the complex gains of each baseline. $\eta_{ij}(t)$ is a noise term resulting from thermal fluctuations in the receivers; $\epsilon_{ij}(t)$ is an offset term. This assumes that the system is linear. π phase switching applied on the first local oscillators is a very efficient method of suppressing the offsets $\epsilon_{ij}(t)$; they are generally negligible and will not be considered any further.

7.1.1 Baseline based vs antenna based gains

Since amplitude and phase distortions have different physical origins it is generally useful to write

$$\mathcal{G}_{ij}(t) = g_i(t) g_j^*(t) = a_i(t) a_j(t) e^{i(\phi_i(t) - \phi_j(t))} \tag{7.2}$$

Here we have split the gains into antenna based factors. This is generally legitimate since the gains represent properties of the data acquisition chains which are in the analogue part of the system. The correlator itself is a digital machine and we assume it is perfectly working (including the clipping correction). This assumption is certainly valid when considering a single frequency and a single instant. When we start averaging in time or frequency, the average of the product may not be the product of averages, and we may have some baseline-based effects.

The baseline-based gains can be determined by observing a point source. This is usually a strong quasar. In that case the true visibilities $V_{ij}(t)$ should all be equal to the quasar flux density S. Then

$$\mathcal{G}_{ij}(t) = \frac{\tilde{V}_{ij}(t)}{S} \tag{7.3}$$

The antenna gains $g_i(t)$ can also be deduced from the non-linear set of equations:

$$g_i(t)g_j^*(t) = \frac{\tilde{V}_{ij}(t)}{S} \tag{7.4}$$

This is a system with N complex unknowns and $N(N-1)/2$ equations. In terms of real quantities there are $N(N-1)$ measured values (amplitudes and phases; there are only $2N-1$ unknowns since one may add a phase factor to all complex gains without affecting the baseline-based complex gains. When N is larger than 2 the system is over determined and may be solved by a method of least squares.

If we note $\tilde{V}_{ij} = \tilde{A}_{ij}e^{i\tilde{\varphi}_{ij}}$, the equations for phases are simply:

$$\phi_i - \phi_j = \tilde{\varphi}_{ij} \tag{7.5}$$

It can be shown that the least-squares solutions (when the same weight is given to all baselines, and if we impose the condition $\sum_{j=1,N} \phi_J = 0$), is given by:

$$\phi_i = \frac{1}{N}\sum_{j\neq i} \tilde{\varphi}_{ij} \tag{7.6}$$

For the amplitudes we can define in order to get a linear system:

$$\gamma_i = \log g_i, \qquad \tilde{\alpha}_{ij} = \log \tilde{A}_{ij} \tag{7.7}$$

$$\gamma_i + \gamma_j = \tilde{\alpha}_{ij} \tag{7.8}$$

This time the least square solution is, when the same weight is given to all baselines:

$$\gamma_i = \frac{1}{N-1}\sum_{j\neq i}\alpha_{ij} - \frac{1}{(N-1)(N-2)}\sum_{j\neq i}\sum_{k\neq i,>j}\alpha_{jk} \tag{7.9}$$

Obviously this antenna gain determination needs at least three antennas. For three antennas it reduces to the obvious result:

$$g_1 = \frac{\tilde{A}_{12}\tilde{A}_{13}}{\tilde{A}_{23}} \tag{7.10}$$

These formulas can be generalized to the cases where the baselines have different weights.

It can be seen in the above formulas that the precision to which the antenna phases and amplitudes is determined is improved by a factor $\sqrt{N}$ over the precision of the measurement of the baseline amplitudes and phases.

7.1.2 Gain corrections

The determination of antenna-based gains (amplitudes and phases) has an obvious advantage: the physical cause of the gain variations are truly antenna-based. One may solve for the gains at the time of the observations, and correct the occurring problems to improve the quality of the data. One may re-point or re-focus the antennas to correct for an amplitude loss, correct for an instrumental delay (affecting the frequency dependence of the phases) ...

7.2 Bandpass calibration

In the previous section we have we have considered a monochromatic system. We have actually finite bandwidths and in principle the gain coefficients are functions of both frequency and time. We thus write:

$$\widetilde{V}_{ij}(\nu,t) = \mathcal{G}_{ij}(\nu,t)V_{ij}(\nu,t) = g_i(\nu,t)g_j^*(\nu,t)V_{ij}(\nu,t) \tag{7.11}$$

If we make the assumption that the passband shape does not change with time, then we should have for each complex baseline-based gain:

$$\mathcal{G}_{ij}(\nu,t) = \mathcal{G}_{\mathrm{B}ij}(\nu)\mathcal{G}_{\mathrm{C}ij}(t) \tag{7.12}$$

The same decomposition can also be done for the antenna-based gains:

$$g_i(\nu,t) = g_{\mathrm{B}i}(\nu)g_{\mathrm{C}i}(t) \tag{7.13}$$

$g_{\mathrm{B}i}(\nu)$ is the antenna complex passband shape, which by convention we normalize so that its integral over the observed bandpass is unity; then $g_{\mathrm{C}i}(t)$ describes the time variation of the complex gains.

Frequency dependence of the gains occurs at several stages in the acquisition chain. In the correlator itself the anti-aliasing filters have to be very steep at the edges of each sub band. A consequence is that the phase slopes can be high there too. Any non-compensated delay offset in the IF can also be seen as a phase linearly dependent on frequency. The attenuation in the cables strongly depends on IF frequency, although this is normally compensated for, to first order, in a well-designed system. The receiver itself has a frequency dependent response both in amplitude and phase, due the IF amplifiers, the frequency dependence on the mixer conversion loss. Antenna chromatism may also be important. Finally the atmosphere itself may have some chromatic behavior, if we operate in the vicinity of a strong line (e.g. O_2 at 118 GHz) or if a weaker line (e.g. O_3) happens to lie in the band.

7.2.1 Bandpass measurement

Bandpass calibration usually relies on observing a very strong source for some time; the bandpass functions are obtained by normalizing the observed visibility spectra by their integral over frequency. It is *a priori* not necessary to observe a point source, as long as its visibility can be assumed to be, on all baselines, independent on frequency in the useful bandwidth. If there is some dependence on frequency, then one should take this into account.

7.2.2 IF passband calibration

In many cases the correlator can be split in several independent sub-bands that are centered to different IF frequencies, and thus observe different frequencies in the sky. In principle they can be treated as different receivers since they have different anti-aliasing filters and different delay offsets, due to different lengths of the connecting cables. Thus they need independent bandpass calibrations, which can be done simultaneously on the same strong source.

At millimeter wavelengths strong sources are scarce, and it is more practical to get a relative calibration of the sub-bands by switching the whole IF inputs to a noise source common to all antennas (Fig. 7.1). The switches are inserted before the IFs are split between sub-bands so that the delay offsets of the sub bands are also calibrated out. This has several advantages: the signal to noise ratio observed by observing the noise source is higher than for an astronomical source since it provides fully correlated signals to the correlator; then such a calibration can be done quite often to suppress any gain drift due to thermal variations in the analogue part of the correlator. Since the sensitivity is high, this calibration is done by baseline, so that any closure errors are taken out.

When such an "IF passband calibration" has been applied in real time, only frequency dependent effects occurring in the signal path before the point where the noise source signal is inserted remain to be calibrated. Since at this point the signal is not yet split between sub-bands, the same passband functions are applicable to all correlator sub-bands.

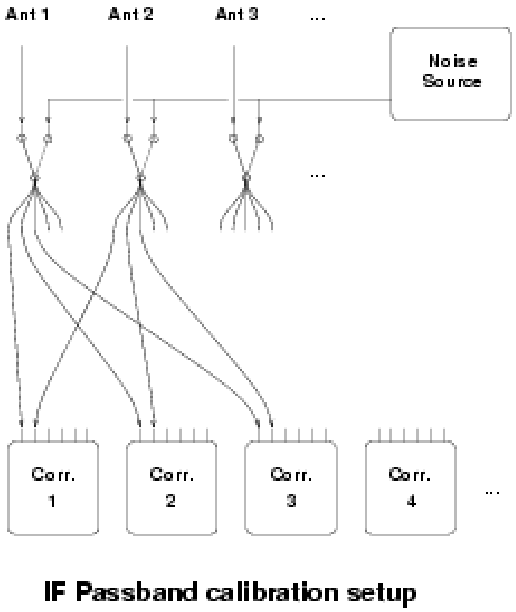

Figure 7.1: IF passband calibration scheme

At Plateau de Bure an "IF passband calibration" is implemented. Of course when the noise source is observed the delay and phase tracking in the last local oscillators (the one in the correlator IF part) are not applied. The precision in phase is $360/\sqrt{\Delta\nu\Delta t} = 0.5°$ at 100 kHz resolution which is sufficient for most projects.

7.2.3 RF bandpass calibration

To actually determine the functions $g_{\mathrm{B}i}(\nu)$ we observe a strong source, with a frequency-independent visibility. The visibilities are

$$\widetilde{V}_{ij}(\nu,t) = g_{\mathrm{B}i}(\nu) g_{\mathrm{B}j}^{*}(\nu) g_{\mathrm{C}i}(t) g_{\mathrm{C}i}^{*}(t) V_{ij}(t) \tag{7.14}$$

Then

$$g_{\mathrm{B}i}(\nu) g_{\mathrm{B}j}^{*}(\nu) = \frac{\widetilde{V}_{ij}(\nu,t)}{\int \widetilde{V}_{ij}(\nu,t) d\nu} \tag{7.15}$$

since the frequency independent factors cancel out in the right-end side. One then averages the measurements on a time long enough to get a decent signal-to-noise ratio. One solves for the antenna-based coefficients in both amplitude and phase; then polynomial amplitude and phase passband curves are fitted to the data.

Applying the passband calibration The passband calibrated visibility data will then be:

$$\widetilde{V}_{\mathrm{C}ij}(\nu,t) = \widetilde{V}_{ij}(\nu,t) / g_{\mathrm{B}i}(\nu) g_{\mathrm{B}j}^{*}(\nu) \tag{7.16}$$

the amplitude and phase of which should be flat functions of frequency.

Accuracy The most important here is the phase precision: it sets the uncertainty for relative positions of spectral features in the map. A rule of thumb is:

$$\Delta\theta/\theta_B = \Delta\phi/360 \tag{7.17}$$

where θ_B is the synthesized beam, and $\Delta\theta$ the relative position uncertainty. The signal to noise ratio on the bandpass calibration should be better than the signal to noise ration of the spectral features observed; otherwise the relative positional accuracy will be limited by the accuracy of the passband calibration.

The amplitude accuracy can be very important too, for instance when one wants to measure a weak line in front of a strong continuum, in particular for a broad line. In that case one needs to measure the passband with an amplitude accuracy better than that is needed on source to get desired signal to noise ratio. Example: we want to measure a line which is 10% of the continuum, with a SNR of 20 on the line strength; then the SNR on the continuum source should be 200, and the SNR on the passband calibration should be at least as good.

7.2.4 Side band calibration

A millimeter-wave interferometer can be used to record separately the signal in both sidebands of the first LO (see Chapter 5). If the first mixer does not attenuate the image side band, then it is useful to average both side bands for increased continuum sensitivity, both for detecting weaker astronomical sources and increasing the SNR for calibration.

However the relative phases of the two sidebands can be arbitrary (particularly at Plateau de Bure where the IF frequency is variable since the LO2 changes in frequency in parallel with the LO1). This relative phase must be calibrated out. This it is done by measuring the phases of the upper and lower sidebands on the passband calibrator observation. These values can be used later to correct each side band phase to compensate for their phase difference.

During the passband calibration one calculates:

$$e^{i\phi_U} = \frac{\int \widetilde{V}_{ij,USB}(\nu,t)d\nu}{|\int \widetilde{V}_{ij,USB}(\nu,t)d\nu|} \tag{7.18}$$

and

$$e^{i\phi_L} = \frac{\int \widetilde{V}_{ij,LSB}(\nu,t)d\nu}{|\int \widetilde{V}_{ij,LSB}(\nu,t)d\nu|} \tag{7.19}$$

Then at any time the double side band visibility is:

$$\widetilde{V}_{ij,DSB}(\nu,t) = e^{-i\phi_U}\widetilde{V}_{ij,USB}(\nu,t) + e^{-i\phi_L}\widetilde{V}_{ij,LSB}(\nu,t) \tag{7.20}$$

As a result the two terms on the right hand side have zero phase at the time of the pass band calibration and they keep the same phase during the whole observing session.

At observing time, offsets on the first and second LOs can be introduced so that both ϕ_U and ϕ_L are very close to zero when a project is done. This actually done at Plateau de Bure, at the same time when the side band gain ratio is measured (see Chapter 10).

7.3 Phase calibration

We now turn to the more difficult problem of correcting for the time dependence of the complex gains, contained in the functions $g_{c\,ij}(t)$. Variations occur in both amplitude and phase. Let us summarize the various effects to be calibrated:

- Interferometer geometry:
  Small errors in the baseline determination will cause slow phase drifts (period 24 hours); unfortunately these errors are dependent on the source direction so they cannot be properly calibrated out by phase referencing on a calibrator, only reduced by a factor of the order of the source to calibrator distance expressed in radians.

- Atmosphere
  The atmosphere introduces phase fluctuations on time scales 1s to a few hours, depending on baseline length and atmospheric conditions (see previous lecture). The effect of fluctuations on short time scales is to cause loss of amplitude by decorrelation, while on the long term the phase fluctuations can be mistaken for structure in the source itself. The critical time there is the time it takes for the projected baseline vector to move by half the diameter of one antenna:

$$\Delta t_C = \frac{D}{2}\frac{86400}{2\pi b} = 224\frac{D}{15}\frac{450}{b} \tag{7.21}$$

  This time ranges from 4 minutes to 1 hour for Plateau de Bure, depending on the baseline length. The phase fluctuations on short time scales may be corrected by applying the radiometric phase correction method, if it is available.

- Antennas
  The antennas may cause variations in amplitude gains due to degradation in the pointing and in the focusing due to thermal deformations in the antennas. These can be corrected to first order by amplitude calibration but it is much better to keep the errors low by proper frequent monitoring of the pointing and focus, since these errors affect differently sources at the center and at the edges of the field. Focus errors also cause strong phase errors due to the additional path length which is twice the sub reflector motion in the axial direction.

- Electronics
  Phase and amplitude drifts in the electronics are kept low by efficient design (see the lectures on receivers and signal and LO transport). The electronics phase drifts are generally slow and of low amplitude, expect for hardware problems.

A detailed analysis can be found in [Lay 1997].

7.3.1 Phase referencing by a nearby point source

This is the standard, traditional way to calibrate the phases with current interferometers. A point source calibrator is typically observed for T_1 (a few minutes) every T_c (20-30 minutes). One fits a gain curve to the data observed on the calibrator, this gain curve is an estimate of the actual gain curve $g_{C\,i}(t)$. This enables removing most long-term phase drifts from the observation of the target source.

Decorrelation It can be shown that the decorrelation factor for a given baseline is approximated by:

$$f \sim 1 - \frac{1}{2}\int_{-log 2.5T_1}^{\infty} \nu P_\phi(\nu) d(\log \nu) \tag{7.22}$$

The decorrelation is fundamentally a baseline-based quantity: it cannot generally be expressed as a product of antenna-based factors. Both the target source and the calibrator are affected, so amplitude referencing will correct for decorrelation. However, the amount of decorrelation will vary from an integration to the next, so that the amplitude uncertainty is increased.

Phase referencing The slow component of $g_{C\,i}(t)$ to be calibrated out is sampled at intervals T_c so that only variations with periods longer than $2T_c$ can be followed. However one fits a slow component into the data points so one is sensitive to errors due to the presence of the fast component: the fast component is *aliased* into a slow component. **It is essential to fit a curve that does not go through the points.**

Fast phase referencing One may reduce T_c as much as possible to remove a larger part of the atmospheric fluctuation spectrum. Time scales of the order of 10s may be used, at the expense of:

- the time efficiency is decreased (relatively more time is spent on the calibrator resulting in a larger overhead)
- caution must be taken than the time T_c may become comparable to the time it takes to water eddies to drift along the apparent distance between source and calibrator.

Water vapor radiometry A radiometry system may be used to monitor the emission of water vapor at a suitable frequency in front of all antennas (dedicated instrument or astronomy receiver; water line or quasi continuum). The fluctuations in the path length can be of a few % of the total path length due to water vapor.

The fluctuations of the water emission are converted into fluctuations of path length by using an atmospheric model (see lecture by M. Bremer, Chapter 9). In principle one could hope to correct for all phase fluctuations this way. However limitations due to receiver stability, to variations in emission from the ground, and to uncertainty in the determination of the emission to path length conversion factor have the consequence that it is not yet possible to consistently correct for the variation of path length between the source and the calibrator.

So far this method at IRAM/Plateau de Bure is used only to correct for on-source fluctuations. Its main effect is to remove the decorrelation effect due to short-term phase fluctuations, improving the precision of amplitude determination.

7.3.2 Phase referencing by a point source in the primary beam

We now consider the simple case where the field contains a strong point source: it can be a continuum source (quasar) or a line source (maser). In that case all phase fluctuations with period longer than $\sim 2T_1$ are removed, where T_1 is the integration time. However statistical errors may be mistaken for true atmospheric phase fluctuations, causing additional decorrelation.

This method gives good results, but for very specific projects which can be observed in very poor atmospheric conditions (e.g. observation of radio emitting quasars, of stars with strong maser lines).

7.3.3 Phase referencing using another band or another frequency

It is generally easier to measure the path lengths fluctuations at a lower frequency (even though the phase scales like frequency), due to both better receiver sensitivity and larger flux of the referencing source. Moreover, in marginal weather conditions, if the rms phase fluctuations at 100 GHz is $\sim 40°$, then at 230 GHz they are of $\sim 100°$, and the phase becomes impossible to track directly due to 2π ambiguities.

If two receivers are available simultaneously, one may subtract to the high frequency phase the phase measured at the low frequency. The atmospheric fluctuations are cancelled and only a slow instrumental drift remains. The gain curve at the high frequency is then determined as the sum of two terms: the low frequency gain curve (including the slow atmospheric terms) plus that slow instrumental drift (which represents any phase fluctuation affecting one of the signal paths of the two receivers).

This method is currently used at Plateau de Bure.