Chapter 4

Two Element Interferometers

Jayaram N. Chengalur

4.1 Introduction

From the van-Cittert Zernike theorem (see Chapter 2) it follows that if one knows the mutual coherence function of the electric field, then the source brightness distribution can be measured[1]. The electric field from the cosmic source is measured using an antenna, which is basically a device for converting the electric field into a voltage that can then be further processed electronically (see Chapter 3). In this chapter we will examine exactly how the mutual coherence function is measured.

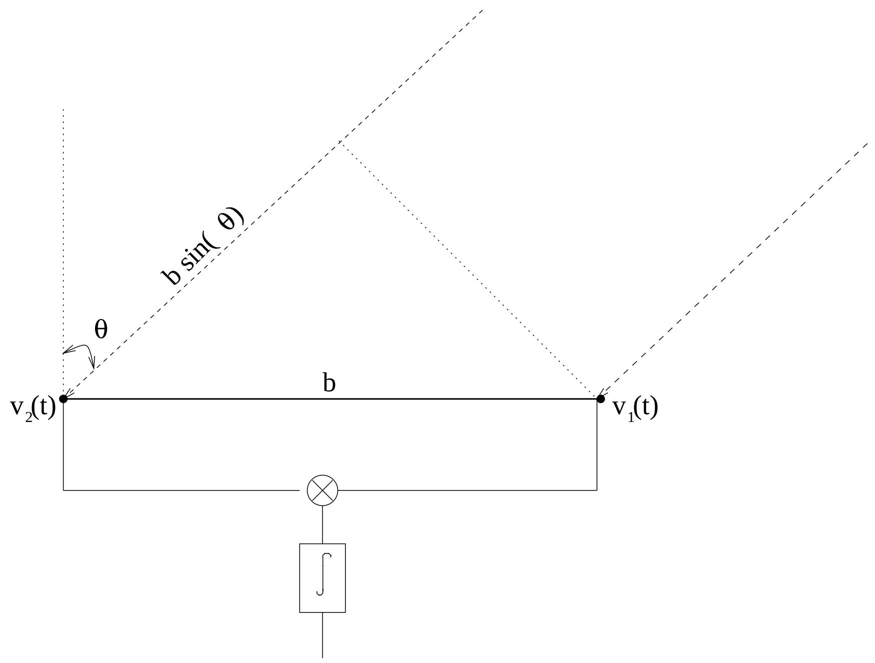

Figure 4.1: A basic two element interferometer

[1]Or in plain english, one make make an image of the source

We start by looking at the relationship between the output of a two element interferometer and the wanted mutual coherence function. Large interferometric arrays can be regarded as collections of two element interferometers, and for this reason it is instructive to understand in detail the working of a two element interferometer.

4.2 A Two Element Interferometer

Consider a two element interferometer shown in Figure 4.1. Two antennas $1, 2$ whose (vector) separation is **b**, are directed towards a point source of flux density S. The angle between the direction to the point source and the normal to the antenna separation vector is θ. The voltages that are produced at the two antennas due to the electric field from this point source are $v_1(t)$ and $v_2(t)$ respectively. These two voltages are multiplied together, and then averaged. Let us start by assuming that the radiation emitted by the source is monochromatic and has frequency ν. Let the voltage at antenna 1 be $v_1(t) = \cos(2\pi\nu t)$. Since the radio waves from the source have to travel an extra distance $b\sin\theta$ to reach antenna 2, the voltage there is delayed by the amount $b\sin\theta/c$. This is called the *geometric delay*, τ_g. The voltage at antenna 2 is hence $v_2(t) = \cos(2\pi\nu(t-\tau_g))$, where we have assumed that the antennas have identical gain. $r(\tau_g)$, the averaged output of the multiplier is hence:

$$
\begin{aligned}
r(\tau_g) &= \frac{1}{\mathrm{T}}\int_{t-T/2}^{t+T/2} \cos(2\pi\nu t)\cos(2\pi\nu(t-\tau_g))dt \\
&= \frac{1}{\mathrm{T}}\int_{t-T/2}^{t+T/2} (\cos(4\pi\nu t - 2\pi\tau_g) + \cos(2\pi\nu\tau_g))dt \\
&= \cos(2\pi\nu\tau_g)
\end{aligned}
\tag{4.2.1}
$$

where we have assumed that the averaging time T is long compared to $1/\nu$. The $\cos(4\pi\nu t)$ factor hence averages out to 0. As the source rises and sets, the angle θ changes. If we assume that the antenna separation vector, (usually called the *baseline vector* or just the *baseline*) is exactly east west, and that the source's declination $\delta_0 = 0$, then $\theta = \Omega_E t$, (where Ω_E is the angular frequency of the earth's rotation) we have:

$$r(\tau_g) = \cos(2\pi\nu \ \times \ b/c \ \times \ \sin(\Omega_E(t - t_z))) \tag{4.2.2}$$

where t_z is the time at which the source is at the zenith. The output $r(\tau_g)$, (also called the *fringe*), hence varies in a quasi-sinusoidal form, with its instantaneous frequency being maximum when the source is at zenith and minimum when the source is either rising or setting (Figure 4.2).

Now if the source's right ascension was known, then one could compute the time at which the source would be at zenith, and hence the time at which the instantaneous fringe frequency would be maximum. If the fringe frequency peaks at some slightly different time, then one knows that assumed right ascension of the source was slightly in error. Thus, in principle at least, from the difference between the actual observed peak time and the expected peak time one could determine the true right ascension of the source. Similarly, if the source were slightly extended, then when the waves from a given point on the source arrive in phase at the two ends of the interferometer, waves arising from adjacent points on the source will arrive slightly out of phase. The observed amplitude of the fringe will hence be less than what would be obtained for a point source of the same total flux. The more extended the source, the lower the fringe amplitude[2]. For a

[2] assuming that the source has a uniform brightness distribution

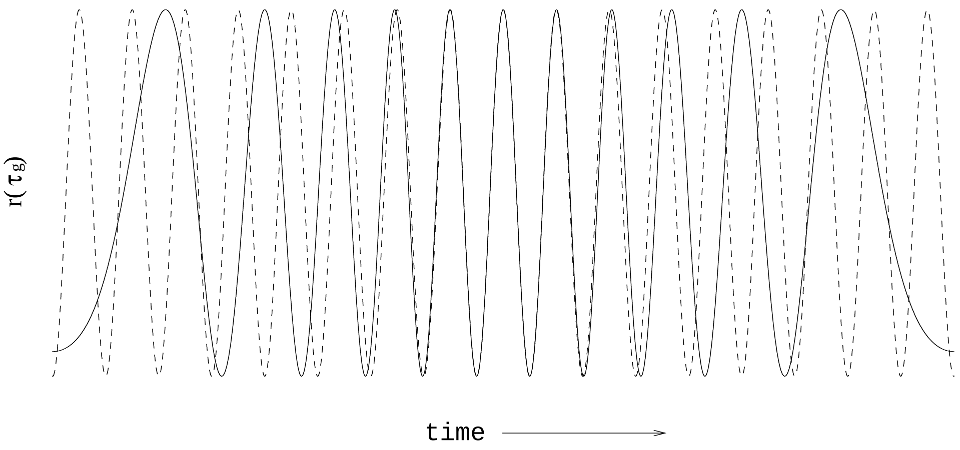

Figure 4.2: The output of a two element interferometer as a function of time. The solid line is the observed qausi-sinosoidal output (the *fringe*), the dotted line is a pure sinusoid whose frequency is equal to the peak instantaneous frequency of the fringe. The instantaneous fringe frequency is maximum when the source is at the zenith (the center of the plot) and is minimum when the source is rising (left extreme) or setting (right extreme).

sufficiently large source with smooth brigtness distribution, the fringe amplitude will be essentially zero[3]. In such circumstances, the interferometer is said to have *resolved out* the source.

Further, two element interferometers cannot distinguish between sources whose sizes are small compared to the fringe spacing, all such sources will appear as point sources. Equivalently when the source size is such that waves from different parts of the source give rise to the same phase lags (within a factor that is small compared to π), then the source will appear as a point source. This condition can be translated into a limit on $\Delta\theta$, the minimum source size that can be resolved by the interferometer, viz.,

$$\pi\nu\Delta\theta b/c \quad \lesssim \quad \pi \qquad \Longrightarrow \qquad \Delta\theta \quad \lesssim \quad \lambda/b$$

i.e., the resolution of a two element interferometer is $\sim \lambda/b$. The longer the baseline, the higher the resolution.

Observations with a two element interferometer hence give one information on both the source position and the source size. Interferometers with different baseline lengths and orientations will place different constraints on the source brightness, and the Fourier transform in the van Cittert-Zernike theorem can be viewed as a way to put all this information together to obtain the correct source brightness distribution.

4.3 Response to Quasi-Monochromatic Radiation

Till now we had assumed that the radiation from the source was monochromatic. Let us now consider the more realistic case of quasi-monochromatic radiation, i.e. the radiation

[3]This is related to the fact that in the double slit experiment, the interference pattern becomes less distinct and then eventually disappears as the source size is increased (see e.g. Born & Wolf, 'Principles of Optics', Sixth Edition, Section 7.3.4). In fact the double slit setup is mathematically equivalent to the two element interferometer, and much of the terminology in radio interferometry is borrowed from earlier optical terminology.

spectrum[4] contains all frequencies in a band $\Delta\nu$ around ν, with $\Delta\nu$ small compared to ν. If the radiation at some frequency ν arrives in phase at the two antennas in the interferometer, the radiation at some adjacent frequencies will arrive out of phase, and if $\Delta\nu$ is large enough, there will be frequencies at which the radiation is actually 180 degrees out of phase. Intuitively hence one would expect that averaging over all these frequencies would decrease the amplitude of the fringe. More rigorously, we have

$$\begin{aligned} r(\tau_g) &= \frac{1}{\Delta\nu}\int_{\nu-\frac{\Delta\nu}{2}}^{\nu+\frac{\Delta\nu}{2}} \cos(2\pi\nu\tau_g)d\nu \qquad (4.3.3)\\ &= \frac{1}{\Delta\nu}Re\left[\int_{\nu-\frac{\Delta\nu}{2}}^{\nu+\frac{\Delta\nu}{2}} e^{i2\pi\nu\tau_g}d\nu\right]\\ &= \cos(2\pi\nu\tau_g)\left[\frac{\sin(\pi\Delta\nu\tau_g)}{\pi\Delta\nu\tau_g}\right] \end{aligned}$$

The quantity in square brackets, the sinc function, decreases rapidly with increasing bandwidth. Hence as one increases the bandwidth that is accepted by the telescope, the fringe amplitude decreases sharply. This is called *fringe washing*. However, since in order to achieve reasonable signal to noise ratio one would require to accept as wide a bandwidth as possible[5], it is necessary to find a way to average over bandwidth without losing fringe amplitude. To understand how this could be done, it is instructive to first look at what the fringe would be for a spatially extended source.

Let the direction vector to some reference point on the source be s_0, and further assume that the source is small that it lies entirely on the tangent plane to the sky at $\mathbf{s_0}$, i.e. that the direction to any point on the source can be written as $\mathrm{s} = \mathbf{s_0} + \boldsymbol{\sigma}$, $\quad \mathbf{s_0}.\boldsymbol{\sigma} = 0$, $\tau_g = \mathbf{s_0}.\mathrm{b}$. Then, from the van Cittert-Zernike theorem we have[6]:

$$\begin{aligned} r(\tau_g) &= Re\left[\int I(\mathbf{s})e^{\frac{-i2\pi\mathbf{s}.\mathbf{b}}{\lambda}}d\mathbf{s}\right]\\ &= Re\left[e^{\frac{-i2\pi\mathbf{s_0}.\mathbf{b}}{\lambda}}\int I(\mathbf{s})e^{\frac{-i2\pi\boldsymbol{\sigma}.\mathbf{b}}{\lambda}}d\mathbf{s}\right]\\ &= |\mathcal{V}|\cos(2\pi\nu\tau_g + \Phi_{\mathcal{V}}) \qquad (4.3.4) \end{aligned}$$

where $\mathcal{V}$, the complex *visibility* is defined as:

$$\mathcal{V} = |\mathcal{V}|e^{-i\Phi_{\mathcal{V}}} = \int I(\mathbf{s})e^{\frac{2\pi\boldsymbol{\sigma}.\mathbf{b}}{\lambda}} \qquad (4.3.5)$$

The information on the source size and structure is contained entirely in $\mathcal{V}$, the factor $\cos(2\pi\nu\tau_g)$ in eqn. (4.3.4) only contains the information that the source rises and sets as the earth rotates. Since this is trivial and uninteresting, it can safely be suppressed. Conceptually, the way one could suppress this information is to introduce along the electrical signal path of antenna 1 an instrumental delay τ_i which is equal to τ_g. Then we will have $r(\tau_g) = |\mathcal{V}|\cos(\Phi_{\mathcal{V}})$, i.e. the fast fringe oscillation has been suppressed. One can then average over frequency and not suffer from fringe washing. Since τ_g changes with time as the source rises and sets, τ_i will also have to be continuously adjusted. This adjustment

[4]Radiation from astrophysical sources is inherently broadband. Radio telescopes however have narrow band filters which accept only a small part of the spectrum of the infalling radiation.

[5]See Chapter 5

[6]apart from some constant factor related to the gain of the antennas which we have been ignoring throughout.

of τ_i is called *delay tracking*. In most existing interferometers however, the process of preventing fringe washing is slightly more complicated than the conceptual scheme described above. The complication arises because delay tracking is usually done digitally in the baseband, i.e. after the whole chain of frequency translation operations described in Chapter 3. The geometric delay is however suffered by the incoming radiation, which is at the RF frequency.

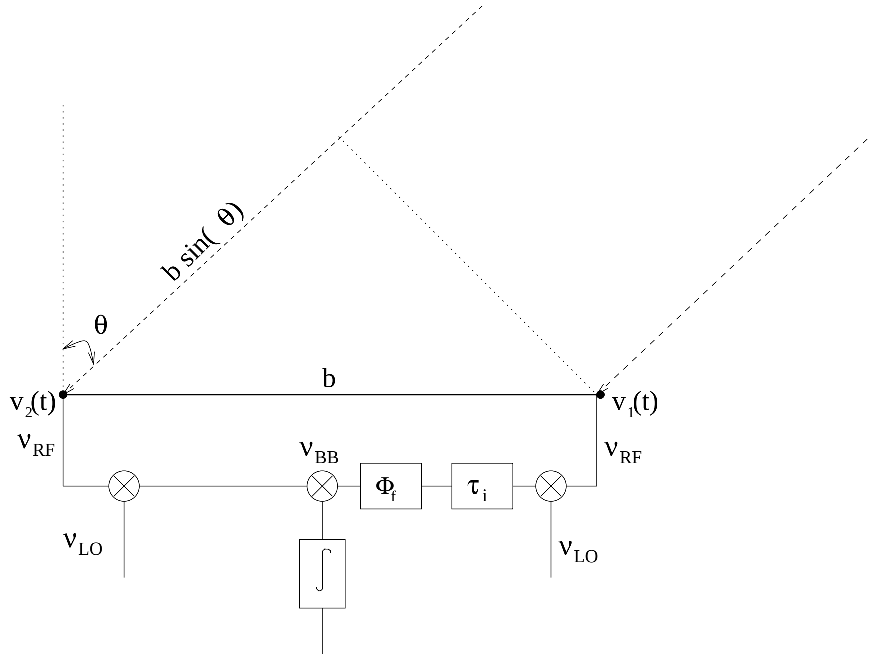

Figure 4.3: A two element interferometer with fringe stopping and delay tracking (see text).

4.4 Two Element Interferometers in Practice

To see this more clearly, let us consider the interferometer shown in Figure 4.3. The signals from antennas $1, 2$ are first converted to a frequency ν_{BB} using a mixer which is fed using a local oscillator of frequency[7] ν_{LO}, i.e. $\nu_{LO} = \nu_{RF} - \nu_{BB}$. Along the signal path for antenna 1 an additional instrumental delay $\tau_i = \tau_g + \Delta\tau$ is introduced, as is also a time varying phase shift Φ_f. The reasons for introducing this phase shift will be clear shortly. Then (see also equations 4.2.1 and 4.3.4) we have:

$$\begin{aligned} r(\tau_g) &= |\mathcal{V}| \left\langle \cos(\Phi_{\mathcal{V}} + 2\pi\nu_{BB}t - 2\pi\nu_{RF}\tau_g)\cos(2\pi\nu_{BB}(t-\tau_i) + \Phi_f)\right\rangle && (4.4.6)\\ &= |\mathcal{V}| cos(\Phi_{\mathcal{V}} + 2\pi(\nu_{RF} - \nu_{BB})\tau_g - \nu_{BB}\Delta\tau - \Phi_f) \\ &= |\mathcal{V}| \cos(\Phi_{\mathcal{V}} + 2\pi\nu_{LO}\tau_g - \nu_{BB}\Delta\tau - \Phi_f) && (4.4.7) \end{aligned}$$

[7]Note that it is important that the phase of the local oscillator signal be identical at the two antennas, i.e. the local oscillator signal has to be distributed in a phase coherent way to both antennas in the interferometer. Chapter 23 explains how this is acheived at the GMRT.

So, in order to compensate for all time varying phase factors, it is not sufficient to have $\tau_i = \tau_g$, one also needs to introduce a time varying phase $\Phi_f = 2\pi\nu_{LO}\tau_g$. This additional correction arises because the delay tracking is done at a frequency different from ν_{RF}. The introduction of the time varying phase is called *fringe stopping*. Fringe stopping can be achieved in a variety of ways. One common practice is to vary the instantaneous phase of the local oscillator signal in arm 1 of the interferometer by the amount Φ_f. Another possibility (which is the approach taken at the GMRT), is to digitally multiply the signal from antenna 1 by a sinusoid with the appropriate instantaneous frequency.

Another consequence of doing delay tracking digitally is that the geometric delay can be quantized only upto a step size which is related to the sampling interval with which the signal was digitized. In general therefore $\Delta\tau$ is not zero, and is called the *fractional sampling time error*. Correction for this error will be discussed in the Chapter 9.

The delay tracking and fringe stopping corrections apply for a specific point in the sky, viz. the position s_0. This point is called the phase tracking center[8]. Signals, such as terrestrial interference, which enter from the far sidelobes of the antennas do not suffer the same geometric delay τ_g as that suffered by the source. Consequently, delay tracking and fringe stopping introduces a rapidly varying change in the phase of these signals. On long baselines, where the fringe rate is rapid, the terrestrial interference could hence get completely decorrelated. While this may appear to be a terrific added bonus, in principle, terrestrial interference is usually so much stronger than the emission from cosmic sources, that even the residual correlation is sufficient to completely swamp out the desired signal.

We end this chapter by re-establishing the connection between what we have just done and the van Cittert-Zernike theorem. The first issue that we have to appreciate is that the van Cittert-Zernike theorem deals with the complex visibility, $\mathcal{V} = |\mathcal{V}|e^{-i\Phi_\mathcal{V}}$. However, the quantity that has been measured is $r(\tau_g) = |\mathcal{V}|\cos(-\Phi_\mathcal{V})$. If one could also measure $|\mathcal{V}|\sin(-\Phi_\mathcal{V})$, then of course one could reconstruct the full complex visibility. This is indeed what is done at interferometers. Conceptually, one has two multipliers instead of the one in Figure 4.3. The second multiplier is fed the same input as that in Figure 4.3, except that an additional phase difference of $\pi/2$ is introduced in each signal path. As can be easily verified, the output of this multiplier is $|\mathcal{V}|\sin(-\Phi_\mathcal{V})$. Such an arrangement of two multipliers is called a *complex correlator*. The two outputs are called the sine and cosine outputs respectively. For quasi-sinsoidal processes, one has to introduce a $\pi/2$ phase difference at each frequency present in the signal. The corresponding transformation is called a *Hilbert transform*[9]. How the complex correlator is achieved at the GMRT is described in Chapter 9. The output of the complex correlator is hence a single component of the Fourier transform of the source brightness distribution[10]. The component measured depends on the antenna separation as viewed from the source, i.e. $(\mathbf{b}.\mathbf{s_0})/\lambda$, which is also called *the projected baseline length*. For a large smooth source, the Fourier transform will be sharply peaked about the origin, and hence the visibility measured on long baselines will be small.

Further Reading

1. Thompson, R. A., Moran, J. M. & Swenson, G. W. Jr., 'Interferometry & Synthesis in Radio Astronomy', Wiley Interscience.

[8]For maximum sensitivity, one would also point the antennas such that their primary beam maxima are also at s_0.

[9]see Chapter 1

[10]This is true only if the antenna dimensions are neglected. Strictly speaking, the measured visibility is an average over the visibilities in the range $\mathbf{b}+\mathbf{a}$ to $\mathbf{b}-\mathbf{a}$ where a is the diameter of the antennas and b is the separation between their midpoints. As will be seen in Chapter 14 the fact that one has information on visibilities on scales smaller than b is useful when attempting to image large regions of the sky.

2. R. A. Perley, F. R. Schwab, & A. H. Bridle, eds., 'Synthesis Imaging in Radio Astronomy', ASP Conf. Series, vol. 6.