Tetsuo Sasao and André B. Fletcher

Introduction to VLBI Systems

Chapter 3

Lecture Notes for KVN Students

Partly Based on Ajou University Lecture Notes

Issued July 14, 2005
revised September 25, 2005
revised October 28, 2005
revised November 11, 2005
revised July 29, 2006
revised October 9, 2009
revised April 20, 2010
(to be further edited)

Radio Interferometry

This chapter describes principles and general features of radio interferometry as the basis of the VLBI.

Contents

1 Fundamentals of Radio Interferometry

1.1 Two Explanations of VLBI

Two quite different explanations on principles of Very Long Baseline Interferometry (VLBI) are given in the literature. The two alternative explanations are illustrated in Figure 1.

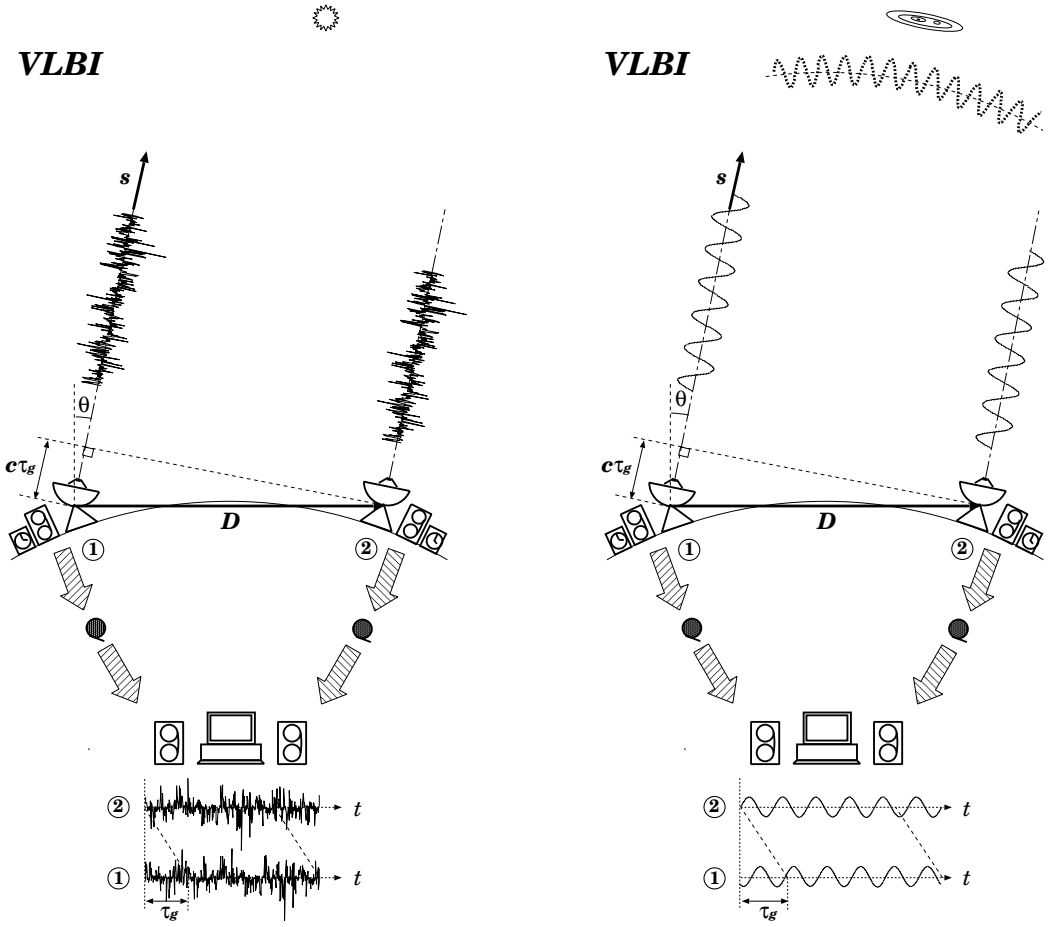

Figure 1: Receptions of noise (left) and monochromatic wave (right) with the VLBI. See Figure 2 for notations $\boldsymbol{s}$, $\boldsymbol{D}$, θ, and τ_g.

1.1.1 VLBI System

The VLBI system itself is described in almost the same way in these two explanations:

Two or more antennas are located at distant stations. They observe the same radio source at the same time. The observed data are recorded on magnetic media, such as magnetic tapes or hard disks, with accurate time marks generated by independent, but highly stable and well synchronized, clocks (or, better to say, frequency standards). The recorded media are sent to a correlation center, where they are played back and mutually multiplied and averaged (integrated) for some duration of time. This "multiplication and integration" procedure is called the "**correlation processing**".

Nowadays, this description of the VLBI system is a little obsolate, since recently observed data are often not recorded but transmitted directly to the correlation center via high–speed communication networks, as we will see in Chapter 4. Nevertheless, the essence, i.e. the correlation of independently time–marked data obtained at distant stations, is kept unchanged.

1.1.2 Geometry of Observation

The geometry of the observation is also presented in a similar way:

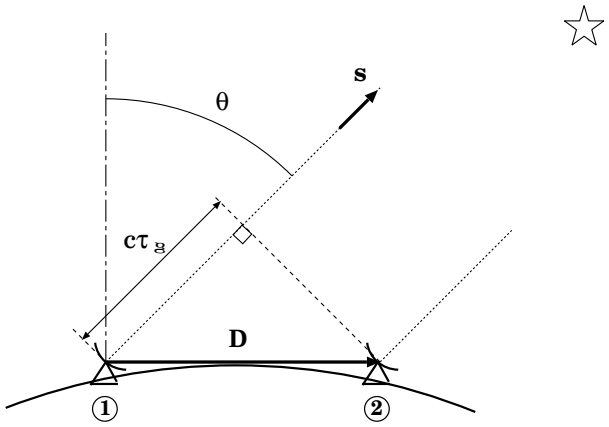

Figure 2: Geometric delay τ_g. $\boldsymbol{D}$: baseline vector connecting reference points of two antennas, $\boldsymbol{s}$: unit vector directed towards a radio source, θ: angle between $\boldsymbol{s}$ and a straight line perpendicular to $\boldsymbol{D}$ in the plane formed by $\boldsymbol{D}$ and $\boldsymbol{s}$, and c: light velocity.

The radio wave from the same radio source must travel slightly further to reach antenna ① (left hand one in Figures 1 and 2) after reaching antenna ② (right hand one), with a small time delay. A

main part of the delay is determined by geometric configuration of the antennas and the radio source, as shown in Figures 1 and 2. Therefore, this is called the "**geometric delay**" and denoted by τ_g in the Figures. In terms of the baseline vector $\boldsymbol{D}$ and the source direction vector $\boldsymbol{s}$, the geometric delay is given by

$$\tau_g = \frac{\boldsymbol{D} \cdot \boldsymbol{s}}{c} = \frac{D \sin\theta}{c}, \tag{1}$$

where c is the light velocity and θ is the angle between $\boldsymbol{s}$ and a line perpendicular to $\boldsymbol{D}$ in the plane formed by $\boldsymbol{D}$ and $\boldsymbol{s}$.

The difference begins with the treatment of the signal from the radio source.

1.1.3 "Geodetic" Explanation — Noise Approach

One explanation, shown in the left panel of Figure 1, which is favored in geodetic VLBI, regards the signal as a **random noise time series** arriving at two antennas with the geometric delay.

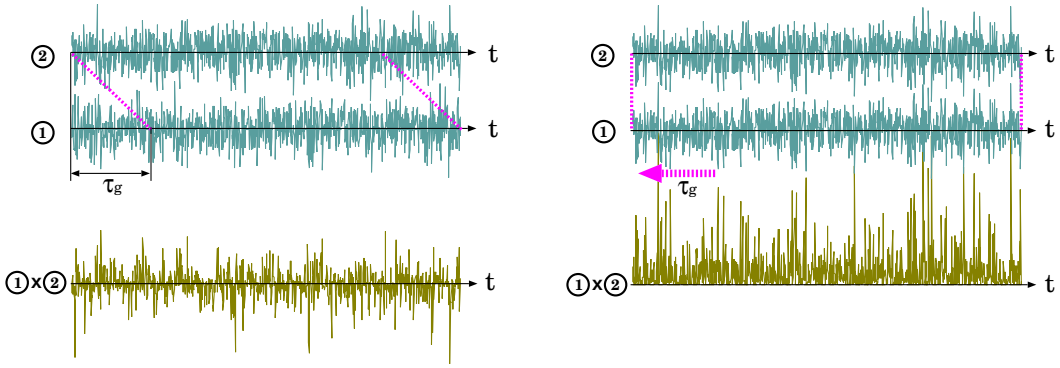

Figure 3: If we have two identical random noise time series one delayed behind another by τ_g (Left, Top and Middle), their product fluctuates around zero (Left, Bottom). Therefore, if we time–average the product we always get nearly zero result. On the other hand, when the delayed series is time–shifted by τ_g, and the two time series get completely aligned (Right, Top and Middle), their product is always non–negative (Right, Bottom). Therefore, if we time–average such a product we get some positive result.

If we simply multiply and average the two played–back data streams recorded at the same time, we must get a nearly zero result in most cases, since we are in effect averaging products of two random noise time series, which is also random noise, with

all possible positive and negative values, as illustrated in the left panel of Figure 3.

But if we shift the playback timing of the record from the antenna ① exactly by the geometric delay τ_g, while keeping the playback timing of the record from the antenna ② unchanged, then the noise patterns from the same source in the two records ① and ② get completely aligned. Therefore, the product of the two time series always gives non–negative values (since plus times plus is plus, and minus times minus is also plus) and the integration yields some finite positive value (right panel of Figure 3).

Thus, we "get the correlation" in the correlation processing (i.e. the multiplication and integration) of the two records. By carefully adjusting the time shift value so that the maximum correlation is obtained, we can precisely determine the geometric delay with an accuracy of 0.1 nsec (10^{-10} sec) or better, which is sufficient to determine the plate movements of the continents with typical speeds of a few centimeter per year.

Mathematically, if signals $v_1(t)$ and $v_2(t)$ received at two antennas ① and ② are mutually time–shifted white noises $n(t-\tau_g)$ and $n(t)$, then multiplication and time averaging give us

$$\overline{v_1(t)\, v_2(t)} \;\propto\; \overline{n(t-\tau_g)\, n(t)} \;\propto\; \delta(\tau_g),$$

as far as we approximate the time average by the statistical ensemble average, where $\overline{(\ )}$ shows the averaging and $\delta(\tau)$ is a delta function of τ.

So, the correlation processing gives non–zero result only when the geometric delay τ_g is well compensated (and therefore $\tau_g = 0$).

1.1.4 "Astrophysical" Explanation — Monochromatic Wave Approach

Another explanation, shown in the right panel of Figure 1, which is favored in astrophysical high–resolution imaging VLBI, regards the signal **as a monochromatic sine wave**.

Two waves from the same source with an angular sky frequency ω arrive at two antennas, giving rise to sinusoidal oscillations with a small time offset τ_g due to the geometric delay. Let the played–back data $v_1(t)$ and $v_2(t)$ from antennas ① and ② be proportional

to $\sin(\omega(t-\tau_g))$ and $\sin(\omega t)$, respectively. Their product is then proportional to

$$v_1(t)\, v_2(t) \propto \sin(\omega(t-\tau_g))\sin(\omega t) = \frac{1}{2}\{\cos(\omega\tau_g) - \cos(2\omega t - \omega\tau_g)\}.$$

It is clear that the contribution of the rapidly oscillating second term in the right hand side of the above equation, at a frequency twice as large as the sky frequency (several billion cycles per second, say), is almost nullified after time averaging (integration) over some duration. Therefore, only the first term, which is proportinal to $\cos(\omega\tau_g)$, is left after the correlation processing, i.e.,

$$\overline{v_1(t)\, v_2(t)} \propto \frac{1}{2}\cos(\omega\tau_g). \tag{2}$$

This term represents a sinusoidal interferometric "**fringe pattern**" on the sky, because the argument $\omega\tau_g$ varies with the source direction $\boldsymbol{s}$, or θ, in the sky, as shown in equation (1). In particular, since $\omega\tau_g = 2\pi c\tau_g/\lambda$, where c is the light velocity, and $\lambda = 2\pi c/\omega$ is the wavelength, the fringe pattern reverses its sign when the path length difference $c\tau_g$ changes by a half wavelength $\lambda/2$, as expected from the standard theory of interferometry. The angular distance corresponding to the separation between two successive peaks of the fringe pattern is called the "**fringe spacing**", as shown in Figure 4.

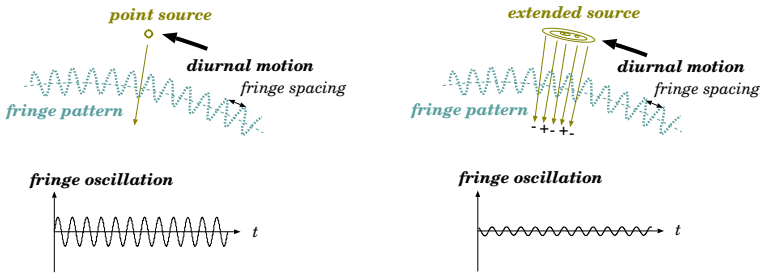

Figure 4: Fringe oscillations of a point source (Left) and of an extended source (Right).

As an observed source moves in the sky due to the diurnal rotation of the Earth, the source moves across the fringe pattern, and the correlator output oscillates in time with the "**fringe frequency**" $\omega\, d\tau_g/dt$ as we expect from equation (2). We call this oscillation as "**fringe oscillation**" and illustrate it in Figure 4.

If we observe a radio source which is sufficiently compact compared with the fringe spacing, the amplitude of the fringe oscillation must be just proportional to the intrinsic strength (the flux density) of the source which is measurable, say, by a single dish radio telescope (left pannel of Figure 4).

If we observe another radio source which is more extended than the fringe spacing, on the other hand, contributions from various elements of the source are mutually compensated in the correlation processing, due to different signs of the fringe pattern over the source extent, as illustrated in the right panel of Figure 4. Therefore, the amplitude of the fringe oscillation is significantly diminished compared with that expected from the intrinsic strength of the source as a whole. The extended source is sort of "resolved" by the fringe pattern. Moreover, the phase of the fringe oscillation is also affected by the source structure.

Thus, the corrrelator output (amplitude and phase of the fringe oscillation) contains information on the structure of the observed radio source. Therefore, by analysing VLBI data obtained with various fringe patterns, we can reconstruct a detailed image of a radio source with surprisingly high angular resolution of 1 milli–arcsecond (about 5 nano–radian) or better, which is sufficient to investigate structures of quasars billions of light years away.

Each of these two explanations, if examined separately, seems clear and internally consistent. But it looks as if they are explaining completely diferent observational technologies, having no common feature at all.

Nevertheless, they are the explanations of the same VLBI, observing the same radio source, with the same antennas, receivers, frequency standards, recorders, and correlators.

Then, how can we understand the two explanations from a unified point of view?

1.1.5 Superposition of Monochromatic Waves

Both of the above two explanations deviate from reality on the same point, but in opposite directions. This point is the spectrum of the received signal (Figure 5).

The noise approach implicitly assumes that the spectrum of the signal is white, i.e. the amplitude of the spectrum is finite, and more or less constant, in a very wide range of frequency. While this assumption may not be too bad for the radio wave propagating in space, it is certainly not valid

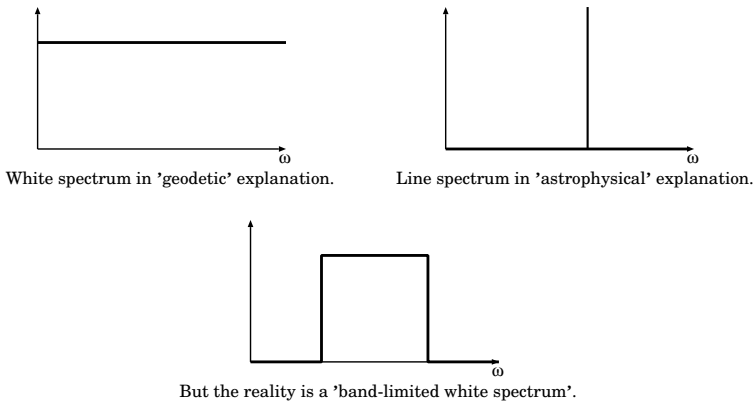

Figure 5: Different source signal spectra assumed in the two explanations.

for the received signal, which must be band–limited due to the frequency characteristics of the optical and receiving systems of element antennas.

The monochromatic–wave approach, on the other hand, assumes an infinitely narrow bandwidth, i.e. a line spectrum, when it talks about a wave having a certain frequency. But this is, of course, far from the reality.

So, what will come out, if we take a more realistic picture, by summing up monochromatic waves with different frequencies, spread within a certain bandwidth? Figure 6 shows an answer.

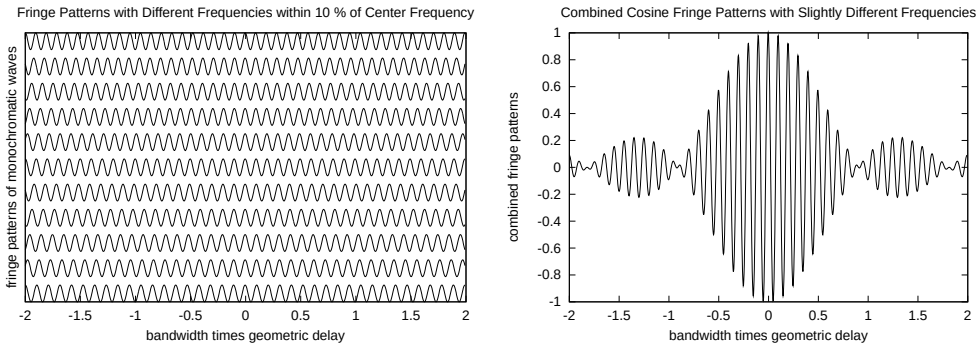

Figure 6: Fringe patterns of 11 monochromatic waves with slightly different frequencies within a bandwidth which is 10 % of the central frequency (left), and their superposition (right).

Here, we summed up 11 fringe patterns of monochromatic waves (left panel of Figure 6), which have slightly different frequencies, distributed at even intervals within a bandwidth B, centered at $\nu_0 = \omega_0/2\pi$, to generate the

superposed pattern shown in the right panel of the same Figure. The lowermost curve, in the left panel of Figure 6, shows the fringe pattern $\cos(\omega_l \tau_g)$ with the lowest angular frequency $\omega_l = 2\pi(\nu_0 - B/2)$, while the uppermost one shows the fringe pattern $\cos(\omega_u \tau_g)$ with the highest angular frequency $\omega_u = 2\pi(\nu_0 + B/2)$.

The horizontal axes of both panels in Figure 6 show the geometric delay τ_g, multiplied by the bandwidth B, within a range of $-2 \leq B\tau_g \leq 2$. We took the center of the horizontal axis at $\tau_g = 0$, since the noise approach predicts that the finite correlation is obtained only when the playback timing of one record is shifted by the geometric delay. This shift is made to align the two records, as if the same wave front is received at the same time by the two antennas. This is obviously equivalent to effectively reducing the geometric delay to zero. Therefore, we assume the simplest case, where the source direction is nearly perpendicular to the baseline, so that $\tau_g \simeq 0$ from the beginning.

We assumed here that the bandwidth is equal to 10 % of the central frequency ($\nu_0 = 10B$).

1.1.6 Fringe Pattern Appears within an Envelope

The right panel of Figure 6 shows a rapid oscillation, enclosed by a more slowly varying envelope.

The rapid oscillation has 10 peaks and valleys, within an interval of $\Delta(B\tau_g) = 1$. The number 10 here is nothing but the ratio ν_0/B. So, this corresponds to the fringe pattern $\cos(\omega_0 \tau_g)$ at the central angular frequency $\omega_0 = 2\pi\nu_0$, as expected in the monochromatic–wave approach.

But the fringe pattern here does not have a constant amplitude. Instead, it is enclosed by an envelope which takes a maximum value at $\tau_g = 0$, when the two time series, obtained from the same source with two antennas, are most coincident with each other. This reminds us of the explanation of the correlation result in the noise approach.

1.1.7 Peak of the Envelope

In order to see the point more clearly, we make our model still closer to an actual continuum spectrum, by increasing the number of monochromatic waves to 100, but keeping the same bandwidth ($B = \nu_0/10$), and show the superposed fringe patterns over a wider range of the horizontal axis: $-20 \leq B\tau_g \leq 20$. The result is given in Figure 7.

Now it is evident that the correlation result of the superposed monochromatic waves has sufficiently large amplitude only within a small range of the

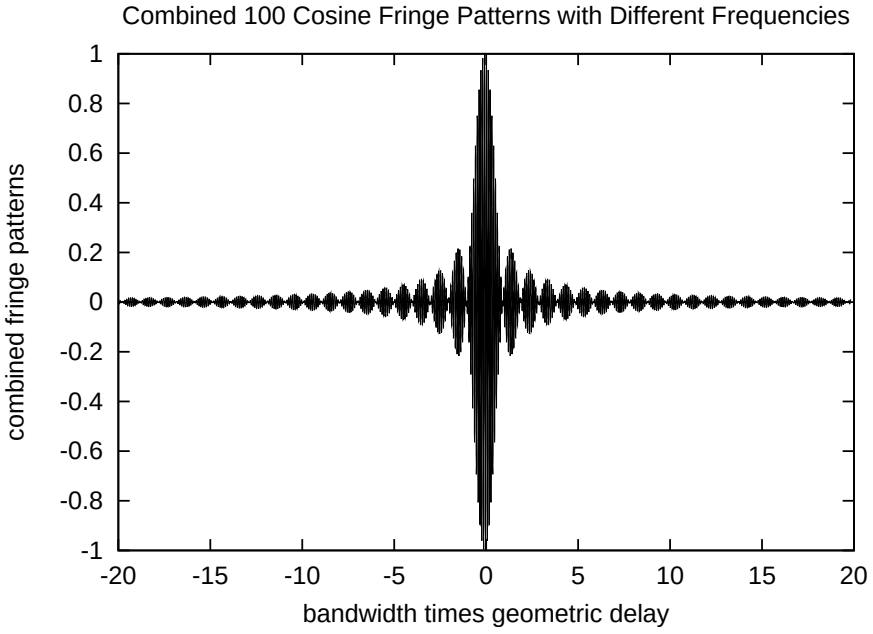

Figure 7: Superposed fringe patterns of 100 monochromatic waves, with slightly different frequencies, contained within a bandwidth equal to 10 % of the central frequency.

geometric delay around $\tau_g = 0$, which is roughly given by $-1/B \leq \tau_g \leq 1/B$. Although there are a number of sidelobes due to the finite bandwidth, the amplitude of these sidelobes rapidly decreases with increasing $|\tau_g|$. Therefore, Figure 7 is actually quite close to what is expected in the noise approach.

This means that we can precisely determine the geometric delay by carefully adjusting the time shift value of the delayed record against another so that the two records are well aligned and the maximum correlation is obtained.

Since the horizontal axis of the Figure 7 stands for $B\tau_g$, the larger the bandwidth, the narrower the envelope is, in terms of the geometric delay τ_g. Therefore, the accuracy of determination of the geometric delay in the geodetic VLBI should be inversely proportional to the bandwidth B.

1.1.8 Fringe Pattern for Radio Source Imaging

Now, as an opposite extreme, let us adopt a narrower bandwidth $B = \nu_0/40$, compared with the central frequency ν_0, and look at a narrower range: $-0.2 \leq B\tau_g \leq 0.2$. The result is given in Figure 8, which clearly shows the fringe pattern $\cos(\omega_0\tau_g)$, at the central angular frequency $\omega_0 = 2\pi\nu_0$ of

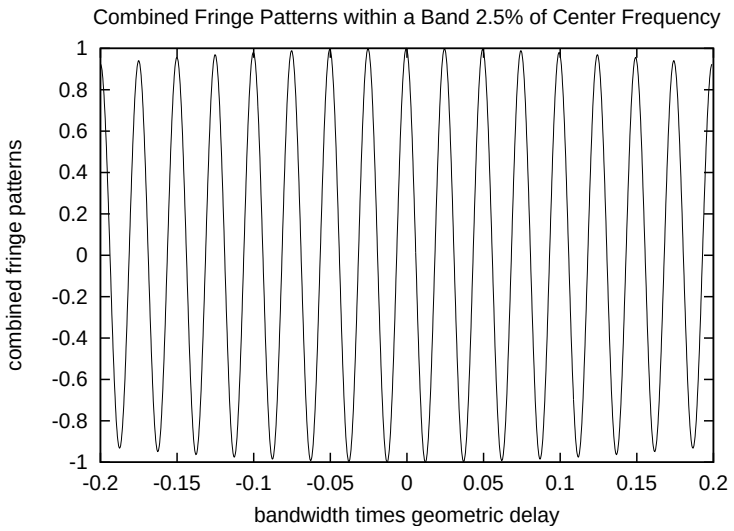

Figure 8: Superposed fringe patterns of monochromatic waves, with slightly different frequencies, contained within a bandwidth equal to 2.5 % of the central frequency.

the band, which is quite similar to the one expected in the monochromatic–wave approach.

Therefore, we can conclude that the two explanations are talking about two extreme cases, corresponding to the very wide and very narrow bandwidths, of a common signal, which is composed of the fringe pattern at the central frequency enclosed by the envelope pattern, whose sharpness is determined by the bandwidth. Geodetic VLBI uses the envelope pattern to determine the peak position, which arises when the signals obtained at two antennas are most coincident with each other, to get a good estimate of the geometric delay τ_g. Source–imaging VLBI uses the fringe pattern, which appears within a limited central range of the envelope pattern, to probe fine structures of astronomical radio sources. Both tasks can be done with the same VLBI telescope.

Although we could obtain, at least qualitatively, a unified view on the apparently quite different two approaches, the above discussion assumed the ensemble of a finite number of monochromatic waves, which are still not very realistic. More rigorous treatment of the signals with band–limited continuum spectra can be obtained in the so–called "white fringe theory" (e.g., Thompson, Moran and Swenson, 2001), which is based on the statistical theory of stationary random processes.

1.2 Elements of Stationary Random Processes

The radio waves coming from astronomical sources are mostly generated by chaotic processes occuring in the source regions. For example, the thermal

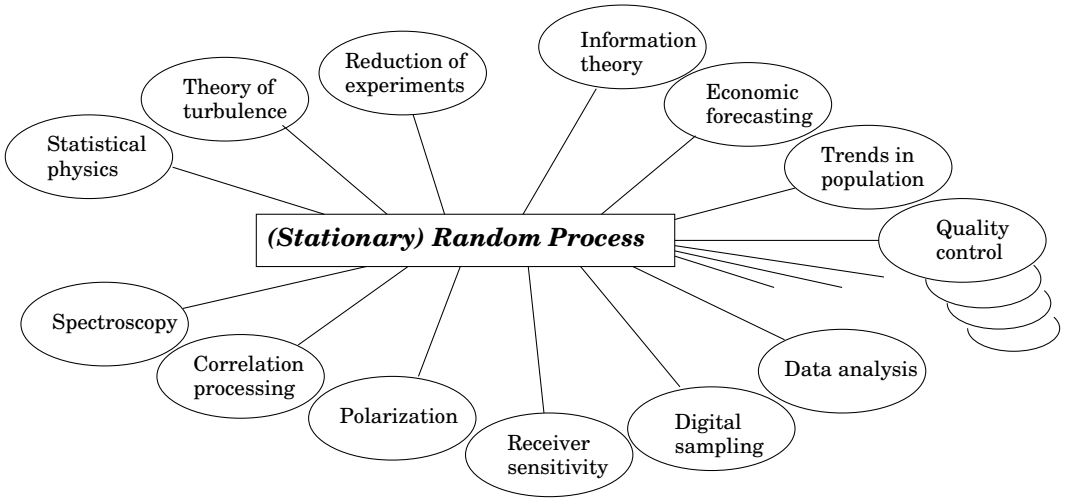

Figure 9: Statistical heory of random processes is a powerful tool for a variety of scientific disciplines.

radiation is caused by thermal random motions of atoms, molecules and free electrons, while the synchrotron radiation emerges from random explosive processes, which accelerate relativistic electrons in magnetic fields. Hence, the electromagnetic fields, or the voltages in the receiving systems, due to the cosmic radio waves, mostly show characteristics of the Gaussian random noise time series, as the "geodetic explanation" assumed. A mathematical tool, which well describes such a random noise time series, is the statistical theory of stationary random processes. The statistical theory of random (or Stochastic) processes has wide applications to many disciplines of radio astronomy, as well as other natural and human sciences, as illustrated in Figure 9.

In the antenna theory, the basic framework was the electromagnetics, and the vector algebra was the main mathematical tool. In the theory of radio interferometry, however, we will no longer newly deal with the electromagnetics. Instead, we will intensively use the theory of stationary random process as a fundamental tool for mathematical development of the interferometry theory.

Therefore, we briefly introduce here basic elements of the theory of stationary random proces to the extent which will be needed in following discus-

sions. For deeper understanding, one can consult with standard textbooks, for example, "Probability, Random Variables, and Stochastic Processes, 2nd Edition" by Athanasios Papoulis (1984) for theory, and "Numerical Recipes" series by Press et al. (1988) for numerical treatment.

1.2.1 Basic Concepts

Random (or Stochastic) Process

A process $x(t)$ is called "random (or Stochastic) process", if it is a function of time t, and, if its value $x(t)$ at any time t is a random variable, i.e., may irregularly vary from trial to trial (see Figure 10).

If we characterize each trial of an experiment by an outcome ζ of the experiment, the random process can be represented as a function of both t and ζ, i.e., as $x(t, \zeta)$.

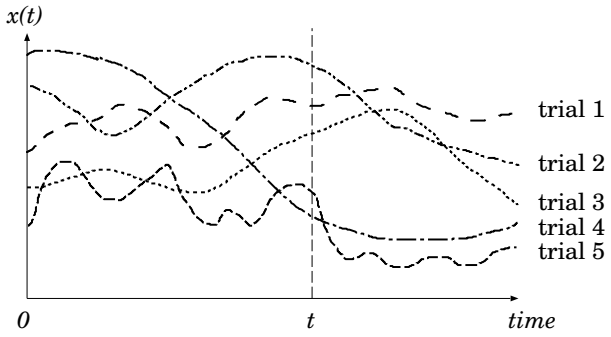

Figure 10: A random process is a function of time whose value at any time t is a random variable.

The random process is a mathematical model of any time–varying and, in general, deterministically unpredictable process. The properties of the random process are usually described in terms of statistical quantities, such as probability distribution, probability density, expectation, correlation, covariance, etc.

Probability Distribution and Probability Density

Let us first consider a real random process $x(t)$.

Let a probability for $x(t)$ at a specific time t not to exceed a certain number x be $P\{x(t) \leq x\}$. Also, let a probability for occurence of multiple

events, $x(t_1)$ not to exceed x_1, $x(t_2)$ not to exceed x_2, $\cdots$, and $x(t_n)$ not to exceed x_n be $P\{x(t_1) \leq x_1,\, x(t_2) \leq x_2,\, \cdots,\, x(t_n) \leq x_n\}$.

Now, the first–order probability distribution $F(x;\, t)$ of the random process $x(t)$ is defined as:

$$F(x;\, t) = P\{x(t) \leq x\}. \tag{3}$$

Likewise, the second–order probability distribution $F(x_1,\, x_2;\, t_1,\, t_2)$ is defined as:

$$F(x_1,\, x_2;\, t_1,\, t_2) = P\{x(t_1) \leq x_1,\, x(t_2) \leq x_2\}, \tag{4}$$

and the n–th–order probability distribution $F(x_1, \cdots,\, x_n;\, t_1, \cdots,\, t_n)$ is defined as:

$$F(x_1, \cdots,\, x_n;\, t_1, \cdots,\, t_n) = P\{x(t_1) \leq x_1, \cdots,\, x(t_n) \leq x_n\}. \tag{5}$$

On the other hand, the first–order probability density $f(x;\, t)$ of the random process $x(t)$ is defined as a derivative of the distribution $F(x;\, t)$ with respect to x:

$$f(x;\, t) = \frac{\partial F(x;\, t)}{\partial x}. \tag{6}$$

Note that $f(x;\, t)$ is really a kind of "density", since, by definition,

$$\frac{\partial F(x;\, t)}{\partial x} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x;\, t) - F(x;\, t)}{\Delta x},$$

therefore, we have

$$\begin{aligned} f(x;\, t) &= \lim_{\Delta x \to 0} \frac{P\{x(t) \leq x + \Delta x\} - P\{x(t) \leq x\}}{\Delta x} \\ &= \lim_{\Delta \to 0} \frac{P\{x < x(t) \leq x + \Delta x\}}{\Delta x}, \end{aligned} \tag{7}$$

which is a probability of $x(t)$ to be "contained in a unit interval of x".

Likewise, the second–order probability density $f(x_1,\, x_2;\, t_1,\, t_2)$ is defined as:

$$f(x_1,\, x_2;\, t_1,\, t_2) = \frac{\partial^2 F(x_1,\, x_2;\, t_1,\, t_2)}{\partial x_1 \partial x_2}, \tag{8}$$

and the n–th–order probability density $f(x_1, \cdots,\, x_n;\, t_1, \cdots,\, t_n)$ is defined as:

$$f(x_1, \cdots,\, x_n;\, t_1, \cdots,\, t_n) = \frac{\partial^n F(x_1, \cdots,\, x_n;\, t_1, \cdots,\, t_n)}{\partial x_1 \cdots \partial x_n}. \tag{9}$$

Generally speaking, if t is continuous, we need infinite number of various orders of probability distributions, in order to properly describe a random process. In many practical cases, especially in cases of stationary random processes, however, it is sufficient to take into account first– and second–order distributions only, as we will see later.

Following general properties are satisfied for probability distributions and densities, as evident from their definitions:

- $F(\infty;\, t) = 1$,
- $F(x_1;\, t_1) = F(x_1,\, \infty;\, t_1,\, t_2)$,
- $f(x;\, t) \geq 0 \quad$ (i.e., $F(x;\, t)$ is a monotonically increasing function of x),
- $\int_{x_1}^{x_2} f(x;\, t)\, dx = P\{x_1 < x(t) \leq x_2\}$,
- $f(x_1;\, t_1) = \int_{-\infty}^{\infty} f(x_1,\, x_2;\, t_1,\, t_2)\, dx_2$,
- $\int_{-\infty}^{\infty} f(x;\, t)\, dx = 1$.

Now, let us consider a more general case, where a random process $z(t)$ is a complex process:

$$z(t) = x(t) + i\, y(t),$$

where a real part $x(t)$ and an imaginary part $y(t)$ are both real random processes, and i is the imaginary unit.

The probability distribution of the complex random process $z(t)$ is defined by the joint probability distribution of $x(t)$ and $y(t)$. Thus, the n–th–order probability distribution is defined as:

$$\begin{aligned} F(z_1, \cdots,\, z_n;\, t_1, \cdots,\, t_n) &= F(x_1, \cdots,\, x_n;\, y_1, \cdots,\, y_n;\, t_1, \cdots,\, t_n) \\ &= P\{x(t_1) \leq x_1, \cdots,\, x(t_n) \leq x_n,\, y(t_1) \leq y_1, \cdots,\, y(t_n) \leq y_n\}, \end{aligned} \tag{10}$$

and the n–th order probability density is defined as:

$$\begin{aligned} f(z_1, \cdots,\, z_n;\, t_1, \cdots,\, t_n) &= f(x_1, \cdots,\, x_n;\, y_1, \cdots,\, y_n;\, t_1, \cdots,\, t_n) \\ &= \frac{\partial^{2n} F(x_1, \cdots,\, x_n;\, y_1, \cdots,\, y_n;\, t_1, \cdots,\, t_n)}{\partial x_1 \cdots \partial x_n\, \partial y_1 \cdots \partial y_n}. \end{aligned} \tag{11}$$

When we have two complex random processes $x(t)$ and $y(t)$:

$$\begin{aligned} x(t) &= x^r(t) + i\, x^i(t), \\ y(t) &= y^r(t) + i\, y^i(t), \end{aligned}$$

where real parts $x^r(t)$ and $y^r(t)$, and imaginary parts $x^i(t)$ and $y^i(t)$, are all real random processes, we introduce joint probability distributions and joint probability densities of the two complex random processes.

For example, the first order joint probability distribution of the complex random processes $x(t)$ and $y(t)$ at times t_1 and t_2, respectively, is

$$\begin{aligned} &F(x;\, y;\, t_1;\, t_2) = F(x^r,\, x^i;\, y^r,\, y^i;\, t_1;\, t_2) \\ &= P\{x^r(t_1) \le x^r;\, x^i(t_1) \le x^i;\, y^r(t_2) \le y^r;\, y^i(t_2) \le y^i\}, \end{aligned} \tag{12}$$

and the corresponding joint probability density is

$$\begin{aligned} &f(x;\, y;\, t_1;\, t_2) = f(x^r,\, x^i;\, y^r,\, y^i;\, t_1;\, t_2) \\ &= \frac{\partial^4 F(x^r,\, x^i;\, y^r,\, y^i;\, t_1;\, t_2)}{\partial x^r\, \partial x^i\, \partial y^r\, \partial y^i}. \end{aligned} \tag{13}$$

Expectation (or Ensemble Average)

Expectation (or ensemble average) $\eta_z(t)$ of a complex random process $z(t) = x(t) + iy(t)$ at time t is defined as:

$$\eta_z(t) = \langle z(t) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + iy)\, f(x;\, y;\, t)\, dxdy, \tag{14}$$

where the symbol $\langle \quad \rangle$, which stands for the expectation, is often denoted also as $E\{ \quad \}$.

Autocorrelation

Autocorrelation $R_{zz}(t_1,\, t_2)$ of a complex random process $z(t) = x(t) + iy(t)$ at times t_1 and t_2 is defined as:

$$\begin{aligned} &R_{zz}(t_1,\, t_2) = \langle z(t_1)\, z^*(t_2) \rangle \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 + iy_1)(x_2 - iy_2)\, f(x_1,\, x_2;\, y_1,\, y_2;\, t_1,\, t_2)\, dx_1 dx_2 dy_1 dy_2, \end{aligned} \tag{15}$$

wehere the symbol $(\quad)^*$ stands for the complex conjugate.

Hereafter, we will usually omit sufficies such as zz, for simplicity, when we express autocorrelations. Thus, autocorrelations $R_{xx}(t_1,\, t_2)$, $R_{yy}(t_1,\, t_2)$, $R_{zz}(t_1,\, t_2)$, and so on, will be all denoted simply as $R(t_1,\, t_2)$, except for cases, when we wish to explicitly specify the random processes under consideration.

Following general properties hold for autocorrelations.

- $R(t_2,\, t_1) = R^*(t_1,\, t_2)$.
- $R(t,\, t) = \langle |\, z(t)\, |^2 \rangle \geq 0, \qquad$ i.e., real and positive.
- Positive definite, namely,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j^* R(t_i,\, t_j) \geq 0 \quad \text{for any numbers } a_i \ (i = 1,\, 2, \cdots,\, n).$$

Proof :

$$0 \leq \langle |\, \sum_{i=1}^{n} a_i z(t_i)\, |^2 \rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j^* \langle z(t_i)\, z^*(t_j) \rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j^* R(t_i,\, t_j).$$

- An inequality:

$$|\, R(t_1,\, t_2)\, |^2 \leq R(t_1,\, t_1)\, R(t_2,\, t_2). \tag{16}$$

Proof :

1. For any complex random variables v and w, we have

$$\langle |\, v\, |\, |\, w\, | \rangle^2 \leq \langle |\, v\, |^2 \rangle \, \langle |\, w\, |^2 \rangle.$$

Proof :
Since we always have

$$\langle (s\, |\, v\, | + |\, w\, |)^2 \rangle = s^2 \langle |\, v\, |^2 \rangle + 2s \langle |\, v\, |\, |\, w\, | \rangle + \langle |\, w\, |^2 \rangle \geq 0,$$

for any real variable s, the discriminant of the above quadratic equation with respect to s must be smaller than or equal to 0, i.e.,

$$\langle |\, v\, |\, |\, w\, | \rangle^2 - \langle |\, v\, |^2 \rangle \, \langle |\, w\, |^2 \rangle \leq 0,$$

and, hence,

$$\langle |\, v\, |\, |\, w\, | \rangle^2 \leq \langle |\, v\, |^2 \rangle \, \langle |\, w\, |^2 \rangle.$$

2. For any complex random variables v and w, we have

$$|\, \langle v\, w \rangle\, | \leq \langle |\, v\, |\, |\, w\, | \rangle.$$

Proof :
Let us first prove a general statement that, for any complex random variable $A = a + ib$, where a and b are real random variables, we have

$$|\, \langle A \rangle\, | \leq \langle |\, A\, | \rangle.$$

Let us denote the probability density of $A = a + ib$ as $f(a,\, b)$. Then $\mid \langle A \rangle \mid$ and $\langle \mid A \mid \rangle$ are expressed as

$$\begin{aligned} \mid \langle A \rangle \mid &= \mid \int \int A\, f(a,\, b)\, dadb \mid \\ &= \lim_{\Delta a \to 0} \lim_{\Delta b \to 0} \mid \sum \sum A\, f(a,\, b) \Delta a \Delta b \mid, \qquad (17) \end{aligned}$$

and

$$\begin{aligned} \langle \mid A \mid \rangle &= \int \int \mid A \mid f(a,\, b)\, dadb \\ &= \lim_{\Delta a \to 0} \lim_{\Delta b \to 0} \sum \sum \mid A \mid f(a,\, b) \Delta a \Delta b \\ &= \lim_{\Delta a \to 0} \lim_{\Delta b \to 0} \sum \sum \mid A\, f(a,\, b) \Delta a \Delta b \mid, \qquad (18) \end{aligned}$$

where we replaced the integrtions by the infinite summations, which are performed in the same way in both equations (17) and (18), and we used a property of the probability density $f(a,\, b)$ in equation (18), that it is always real and greater than or equal to zero.

Now, for any complex numbers B and C, we have

$$\mid B + C \mid \leq \mid B \mid + \mid C \mid,$$

since

$$\begin{aligned} \mid B + C \mid &= \sqrt{(B + C)(B + C)^*} = \sqrt{\mid B \mid^2 + B^* C + B C^* + \mid C \mid^2} \\ &= \sqrt{\mid B \mid^2 + 2 \mid B \mid \mid C \mid \cos \Phi + \mid C \mid^2}, \end{aligned}$$

where we introduced an angle Φ satisfying

$$B^* C = \mid B \mid \mid C \mid e^{i\Phi},$$

is always smaller than $\mid B \mid + \mid C \mid$, because $\cos \Phi \leq 1$. This relation is easily extended to the sum of arbitrary number n of complex numbers $B_1,\, B,\, \cdots,\, B_n$, i.e.,

$$\mid \sum_{i=1}^{n} B_i \mid \leq \sum_{i=1}^{n} \mid B_i \mid,$$

because

$$\begin{aligned} \mid \sum_{i=1}^{n} B_i \mid &= \mid B_1 + \sum_{i=2}^{n} B_i \mid \leq \mid B_1 \mid + \mid \sum_{i=2}^{n} B_i \mid = \mid B_1 \mid + \mid B_2 + \sum_{i=3}^{n} B_i \mid \\ &\leq \mid B_1 \mid + \mid B_2 \mid + \mid \sum_{i=3}^{n} B_i \mid = \cdots \leq \mid B_1 \mid + \mid B_2 \mid + \cdots + \mid B_n \mid . \end{aligned}$$

Applying the above relation to the summations of equations (17) and (18), we confirm that $|\langle A\rangle| \leq \langle|A|\rangle$.
This implies that $|\langle v\,w\rangle| \leq \langle|v||w|\rangle$, since

$$|v\,w| = \sqrt{vwv^*w^*} = \sqrt{vv^*}\sqrt{ww^*} = |v||w|.$$

3. From 1. and 2. above, we obtain

$$|\langle v\,w\rangle|^2 \leq \langle|v||w|\rangle^2 \leq \langle|v|^2\rangle\langle|w|^2\rangle,$$

i.e.,

$$|\langle v\,w\rangle|^2 \leq \langle|v|^2\rangle\langle|w|^2\rangle. \tag{19}$$

If we adopt here $v = z(t_1)$ and $w = z^*(t_2)$, then we prove that

$$|R(t_1, t_2)|^2 \leq R(t_1, t_1)\,R(t_2, t_2).$$

Autocovariance

Autocovariance $C(t_1, t_2)$ of a complex random process $z(t)$ at times t_1 and t_2 is defined as:

$$C(t_1, t_2) = R(t_1, t_2) - \eta(t_1)\,\eta^*(t_2), \tag{20}$$

where $\eta(t) \equiv \eta_z(t) = \langle z(t)\rangle$ is the expectation of $z(t)$ at time t.

The autocovariance of $z(t)$ is equal to the autocorrelation of $\tilde{z}(t) = z(t) - \eta(t)$, i.e.,

$$C(t_1, t_2) = R_{\tilde{z}\tilde{z}} = \langle\tilde{z}(t_1)\,\tilde{z}^*(t_2)\rangle. \tag{21}$$

In fact,

$$\begin{aligned}\langle\tilde{z}(t_1)\,\tilde{z}^*(t_2)\rangle &= \langle[z(t_1)\,z^*(t_2) - z(t_1)\,\eta^*(t_2) - \eta(t_1)\,z^*(t_2) + \eta(t_1)\,\eta^*(t_2)]\rangle \\ &= R(t_1, t_2) - \eta(t_1)\eta^*(t_2).\end{aligned}$$

Correlation Coefficient

Correlation coefficient $r(t_1, t_2)$ of a complex random process $z(t)$ at times t_1 and t_2 is defined as:

$$r(t_1, t_2) = \frac{C(t_1, t_2)}{\sqrt{C(t_1, t_1)\,C(t_2, t_2)}}. \tag{22}$$

It is evident that

- $r(t,\, t) = 1$.

Also, the absolute value of the correlation coefficient is always smaller than or equal to 1:

- $\mid r(t_1,\, t_2) \mid \leq 1$,

since from equations (16) and (21), we have

$$\mid C(t_1,\, t_2) \mid^2 \leq C(t_1,\, t_1)\, C(t_2,\, t_2).$$

Cross–Correlation

Cross–correlation $R_{xy}(t_1,\, t_2)$ of two complex random processes $x(t) = x^r(t) + i\, x^i(t)$ and $y(t) = y^r(t) + i\, y^i(t)$ at times t_1 and t_2, respectively, is defined as:

$$R_{xy}(t_1,\, t_2) = \langle x(t_1)\, y^*(t_2) \rangle =$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x^r + ix^i)\, (y^r - iy^i)\, f(x^r,\, x^i;\, y^r,\, y^i;\, t_1;\, t_2) dx^r\, dx^i\, dy^r\, dy^i, \quad (23)$$

using the joint probability density of $x(t)$ and $y(t)$, given in equation (13).

Following properties hold for cross–correlations.

- $R_{xy}(t_2,\, t_1) = R^*_{yx}(t_1,\, t_2)$.
  This is evident from the above definition.

- $\mid R_{xy}(t_1,\, t_2) \mid^2 \leq\ R_{xx}(t_1,\, t_1)\, R_{yy}(t_2,\, t_2)$.
  We can readily prove this by adopting $v = x(t_1)$ and $w = y^*(t_2)$ in equation (19).

Cross–Covariance

Cross–covariance $C_{xy}(t_1,\, t_2)$ of two complex random processes $x(t)$ and $y(t)$ at times t_1 and t_2 is defined as:

$$C_{xy}(t_1,\, t_2) = R_{xy}(t_1,\, t_2) - \eta_x(t_1)\, \eta^*_y(t_2). \quad (24)$$

Following properties hold for cross-covariances.

- The cross–covariance of $x(t)$ and $y(t)$ is equal to the cross–correlation of $\tilde{x}(t) = x(t) - \eta_x(t)$ and $\tilde{y}(t) = y(t) - \eta_y(t)$, i.e.,

$$C_{xy}(t_1,\, t_2) = R_{\tilde{x}\tilde{y}}(t_1,\, t_2) = \langle \tilde{x}(t_1)\, \tilde{y}^*(t_2) \rangle.$$

- $\mid C_{xy}(t_1,\, t_2) \mid^2 \leq\ C_{xx}(t_1,\, t_1)\, C_{yy}(t_2,\, t_2)$.

Cross–Correlation Coefficient

Cross–correlation coefficient $r_{xy}(t_1,\, t_2)$ of two complex random processes $x(t)$ and $y(t)$ at times t_1 and t_2 is defined as:

$$r_{xy}(t_1,\, t_2) = \frac{C_{xy}(t_1,\, t_2)}{\sqrt{C_{xx}(t_1,\, t_1)\, C_{yy}(t_2,\, t_2)}}. \tag{25}$$

It is evident that for any cross–correlation coefficient we always have

- $\mid r_{xy}(t_1,\, t_2) \mid \leq 1$.

Normal (Gaussian) Process

As an example of the probability density introduced in equations (6), (8), and (9), we consider here probability desnity of a particularly important class of random process, namely normal (or Gaussian) process, which is known to be a good model of signals from astronomical radio sources, as well as of noises produced in antenna–receiving systems or in environments.

Real random process $x(t)$, with expectation $\eta(t)$ and autocovariance $C(t_1,\, t_2)$, is called "normal (or Gaussian) process", if, at any times $t_1,\, t_2,\, \cdots,\, t_n$ for any n, random variables $x(t_1),\, x(t_2),\, \cdots,\, x(t_n)$ are jointly normal (or Gaussian), i.e., they are characterized by following probabilty densities.

- First–order Gaussian probability density:

$$f(x_1;\, t_1) = \frac{1}{\sigma_1\sqrt{2\pi}}\, e^{-\frac{(x_1-\eta_1)^2}{2\sigma_1^2}}, \tag{26}$$

  where $\eta_1 \equiv \eta(t_1)$ and $\sigma_1^2 \equiv C(t_1,\, t_1) = \langle [x(t_1) - \eta(t_1)]^2 \rangle$ are expectation and so–called dispersion, respectively, of $x(t)$ at time t_1.

- Second–order Gaussian probability density:

$$\begin{aligned} &f(x_1,\, x_2;\, t_1,\, t_2) \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\, e^{-\frac{1}{2(1-r^2)}\left(\frac{(x_1-\eta_1)^2}{\sigma_1^2} - 2r\frac{(x_1-\eta_1)(x_2-\eta_2)}{\sigma_1\sigma_2} + \frac{(x_2-\eta_2)^2}{\sigma_2^2}\right)}, \end{aligned} \tag{27}$$

  where we introduced notations: $\eta_1 \equiv \eta(t_1)$, $\eta_2 \equiv \eta(t_2)$, $\sigma_1^2 \equiv C(t_1,\, t_1)$, $\sigma_2^2 \equiv C(t_2,\, t_2)$, and correlation coefficient:

$$r \equiv \frac{C(t_1,\, t_2)}{\sqrt{C(t_1,\, t_1)\, C(t_2,\, t_2)}}.$$

- n-th–order Gaussian probability density:

$$f(x_1, \cdots, x_n; t_1, \cdots, t_n) = \frac{1}{\sqrt{(2\pi)^n \Delta}} e^{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i-\eta_i)\, C_{ij}^{-1}\,(x_j-\eta_j)}, \tag{28}$$

  where $\eta_i \equiv \eta(t_i)$, $\eta_j \equiv \eta(t_j)$, $C_{ij} \equiv C(t_i, t_j)$ is autocovariance matrix, C_{ij}^{-1} is its inverse, and $\Delta \equiv \det\{C_{ij}\}$ is its determinant.

  When $n = 1$ and $n = 2$, equation (28) is reduced to equations (26) and (27), respectively, implying that the first– and second–order Gaussian probability densities given in equations (26) and (27), respectively, are special cases of the more general expression of the n-th–order Gaussian (or normal) probability density given in equation (28).

Likewise, we can conceive a number of normal (or Gaussian) processes which are jointly normal with each other.

- Two real normal processes $x(t)$ and $y(t)$ are called "jointly normal (or Gaussian) processes", if, at any times t_1 and t_2, random variables $x(t_1)$ and $y(t_2)$ are jointly normal, i.e., they are characterized by following joint probabilty density:

$$f(x, y; t_1, t_2) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r_{xy}^2}} e^{-\frac{1}{2(1-r_{xy}^2)}\left(\frac{(x-\eta_x)^2}{\sigma_x^2} - 2\, r_{xy}\frac{(x-\eta_x)(y-\eta_y)}{\sigma_x\sigma_y} + \frac{(y-\eta_y)^2}{\sigma_y^2}\right)}, \tag{29}$$

  where we introduced expectations of $x(t_1)$ and $y(t_2)$: $\eta_x \equiv \eta_x(t_1)$ and $\eta_y \equiv \eta_y(t_2)$, dispersions of $x(t_1)$ and $y(t_2)$: $\sigma_x^2 \equiv C_{xx}(t_1, t_1)$, $\sigma_y^2 \equiv C_{yy}(t_2, t_2)$, and cross–correlation coefficient:

$$r_{xy} \equiv \frac{C_{xy}(t_1, t_2)}{\sqrt{C_{xx}(t_1, t_1)\, C_{yy}(t_2, t_2)}}.$$

  Here $C_{xx}(t_1, t_1)$, $C_{yy}(t_2, t_2)$ and $C_{xy}(t_1, t_2)$ are autocovariances and cross–covariance of $x(t_1)$ and $y(t_2)$, correspondingly.

- An arbitrary number m of real normal processes $x_{(1)}(t)$, $x_{(2)}(t)$, $\cdots$, $x_{(m)}(t)$ are called "jointly normal (or Gaussian) processes", if, at any times $t_1, t_2, \cdots, t_m$, random variables $x_{(1)}(t_1)$, $x_{(2)}(t_2)$, $\cdots$, $x_{(m)}(t_m)$ are

jointly normal, i.e., they are characterized by following joint probabilty density:

$$f(x_{(1)}, \cdots, x_{(m)}; t_1, \cdots, t_m) = \frac{1}{\sqrt{(2\pi)^m \Delta}} e^{-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(x_{(i)}-\eta_{(i)})\, C^{-1}_{(i)(j)}\, (x_{(j)}-\eta_{(j)})}, \qquad (30)$$

where $\eta_{(i)} = \langle x_{(i)}(t_i)\rangle$ and $\eta_{(j)} = \langle x_{(j)}(t_j)\rangle$ are expectations of $x_{(i)}(t)$ at $t = t_i$ and $x_{(j)}(t)$ at $t = t_j$, $C_{(i)(j)} \equiv C_{x_{(i)}x_{(j)}}(t_i, t_j)$ is cross–covariance matrix, $C^{-1}_{(i)(j)}$ is inverse matrix of $C_{(i)(j)}$, and $\Delta \equiv \det\{C_{(i)(j)}\}$ is determinant of $C_{(i)(j)}$.

Of course, equation (29) is a special case of equation (30) with $m = 2$.

- In the above statement, some random variables among $x_{(1)}(t_1)$, $x_{(2)}(t_2)$, $\cdots$, $x_{(m)}(t_m)$ could be values of the same normal process taken at different times. In such a case, some elements of matrix $C_{(i)(j)}$ are auto-covariances, rather than cross–covariances.

  In this sence, joint probability densities of a single normal (or Gaussian) process given in equations (26), (27), and (28) can be regarded as special cases of the joint probability density of the jointly normal (or Gaussian) processes given in equation (30).

It is not difficult to confirm that the joint probability densities of jointly normal (or Gaussian) processes given in equations (26) – (30) are consistent with general properties of joint probability densities, as well as with definitions of the expectation and the covariances, as explained in standard textbooks. For this purpose, we can use well–known integration formulae:

$$\int_{-\infty}^{\infty} e^{-a\,x^2}\, dx = \sqrt{\frac{\pi}{a}}, \qquad (31)$$

$$\int_{-\infty}^{\infty} x\, e^{-a\,x^2}\, dx = 0, \qquad (32)$$

$$\int_{-\infty}^{\infty} x^2\, e^{-a\,x^2}\, dx = \frac{1}{2}\sqrt{\frac{\pi}{a^3}}, \qquad (33)$$

for $a > 0$.

For example, if we take the form of the joint normal (or Gaussian) probability density given in equation (29), we can confirm following formulae.

- $\int\limits_{-\infty}^{\infty} f(x,\, y;\, t_1,\, t_2)\, dy = f(x\,;\, t_1)$:

$$\begin{aligned}
&\int\limits_{-\infty}^{\infty} f(x,\, y;\, t_1,\, t_2)\, dy \\
&= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r_{xy}^2}} \int\limits_{-\infty}^{\infty} e^{-\frac{1}{2(1-r_{xy}^2)}\left(\frac{(x-\eta_x)^2}{\sigma_x^2} - 2\,r_{xy}\frac{(x-\eta_x)(y-\eta_y)}{\sigma_x\sigma_y} + \frac{(y-\eta_y)^2}{\sigma_y^2}\right)} dy \\
&= \frac{1}{2\pi\sigma_x\sqrt{1-r_{xy}^2}} \int\limits_{-\infty}^{\infty} e^{-\frac{1}{2(1-r_{xy}^2)}\left(\frac{(x-\eta_x)^2}{\sigma_x^2} - 2\,r_{xy}\frac{(x-\eta_x)}{\sigma_x}\,y' + y'^2\right)} dy' \\
&= \frac{1}{2\pi\sigma_x\sqrt{1-r_{xy}^2}} \int\limits_{-\infty}^{\infty} e^{-\frac{1}{2(1-r_{xy}^2)}\left[(1-r_{xy}^2)\,\frac{(x-\eta_x)^2}{\sigma_x^2} + \left(y' - r_{xy}\frac{x-\eta_x}{\sigma_x}\right)^2\right]} dy' \\
&= \frac{1}{2\pi\sigma_x\sqrt{1-r_{xy}^2}}\, e^{-\frac{(x-\eta_x)^2}{2\,\sigma_x^2}} \int\limits_{-\infty}^{\infty} e^{-\frac{y''^2}{2\,(1-r_{xy}^2)}}\, dy'' \\
&= \frac{1}{\sigma_x\,\sqrt{2\,\pi}}\, e^{-\frac{(x-\eta_x)^2}{2\,\sigma_x^2}} \;=\; f(x\,;\, t_1), \qquad (34)
\end{aligned}$$

in view of equation (31), where we introduced variable transformations:

$$y' = \frac{y-\eta_y}{\sigma_y}, \quad y'' = y' - r_{xy}\,\frac{x-\eta_x}{\sigma_x}.$$

- Expectation:

$$\begin{aligned}
\langle x\rangle &= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} x\, f(x,\, y;\, t_1,\, t_2)\, dx\, dy \;=\; \int\limits_{-\infty}^{\infty} x \int\limits_{-\infty}^{\infty} f(x,\, y;\, t_1,\, t_2)\, dy\, dx \\
&= \int\limits_{-\infty}^{\infty} x f(x\,;\, t_1)\, dx \;=\; \frac{1}{\sigma_x\,\sqrt{2\,\pi}} \int\limits_{-\infty}^{\infty} x\, e^{-\frac{(x-\eta_x)^2}{2\,\sigma_x^2}}\, dx \\
&= \frac{1}{\sigma_x\,\sqrt{2\,\pi}} \int\limits_{-\infty}^{\infty} [(x-\eta_x) + \eta_x]\, e^{-\frac{(x-\eta_x)^2}{2\,\sigma_x^2}}\, dx \\
&= \frac{\eta_x}{\sigma_x\,\sqrt{2\,\pi}} \int\limits_{-\infty}^{\infty} e^{-\frac{(x-\eta_x)^2}{2\,\sigma_x^2}}\, dx \;=\; \eta_x, \qquad (35)
\end{aligned}$$

in view of equations (31) and (32).

The same argument also leads to

$$\langle y\rangle = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} y\, f(x,\, y;\, t_1,\, t_2)\, dx\, dy \;=\; \eta_y.$$

- Covariance:

$$
\begin{aligned}
&\langle (x-\eta_x)\,(y-\eta_y)\rangle \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\eta_x)\,(y-\eta_y)\,f(x,\,y;\,t_1,\,t_2)\,dx\,dy \\
&= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r_{xy}^2}} \\
&\times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\eta_x)\,(y-\eta_y)\,e^{-\frac{1}{2(1-r_{xy}^2)}\left(\frac{(x-\eta_x)^2}{\sigma_x^2}-2\,r_{xy}\frac{(x-\eta_x)(y-\eta_y)}{\sigma_x\sigma_y}+\frac{(y-\eta_y)^2}{\sigma_y^2}\right)}\,dx\,dy \\
&= \frac{\sigma_x\,\sigma_y}{2\pi\sqrt{1-r_{xy}^2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x'\,y'\,e^{-\frac{1}{2(1-r_{xy}^2)}(x'^2-2r_{xy}\,x'\,y'+y'^2)}\,dx'\,dy' \\
&= \frac{\sigma_x\,\sigma_y}{2\pi\sqrt{1-r_{xy}^2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x''+r_{xy}\,y'')\,y''\,e^{-\frac{1}{2(1-r_{xy}^2)}[x''^2+(1-r_{xy}^2)\,y''^2]}\,dx''\,dy'' \\
&= \frac{\sigma_x\,\sigma_y\,r_{xy}}{2\pi\sqrt{1-r_{xy}^2}} \int_{-\infty}^{\infty} e^{-\frac{x''^2}{2(1-r_{xy}^2)}}\,dx'' \int_{-\infty}^{\infty} y''^2\,e^{-\frac{y''^2}{2}}\,dy'' \\
&= \frac{\sigma_x\,\sigma_y\,r_{xy}}{2\pi\sqrt{1-r_{xy}^2}}\,\sqrt{2\pi(1-r_{xy}^2)}\,\sqrt{2\pi} \;=\; r_{xy}\,\sigma_x\,\sigma_y \;=\; C_{xy}(t_1,\,t_2), \qquad (36)
\end{aligned}
$$

  in view of equations (31), (32), and (33), where we introduced variable transformations:

$$
\begin{aligned}
x' &= \frac{x-\eta_x}{\sigma_x}, \qquad y' = \frac{y-\eta_y}{\sigma_y}, \\
x'' &= x' - r_{xy}\,y', \qquad y'' = y'.
\end{aligned}
$$

In following discussions, we will mainly use general properties of expectations and correlations, without specifying explicit forms of probability densities. However, when necessary, we will assume jointly normal (or Gaussian) processes, and explicitly use expressions of the normal probability density given in equations (26) – (30).

1.2.2 Random Processes in Linear Systems

Definition of Linear Systems

Let us consider a system of two complex functions $x(t)$ and $y(t)$ of time t, which are related with each other by an operator L:

$$y(t) = L[x(t)], \qquad (37)$$

where $x(t)$ is called "input" and $y(t)$ is called "output" of the operator L.

Such a system is called "linear system", if the operator L satisfies

$$L[a_1\, x_1(t) + a_2\, x_2(t)] = a_1 L[x_1(t)] + a_2 L[x_2(t)], \tag{38}$$

for any complex coefficients a_1, a_2 and for any functions $x_1(t)$, $x_2(t)$.

The linear system is also called as "linear filter", which linearly "filters" the input $x(t)$ to yield the output $y(t)$.

Impulse Response

If the input of an operator L is a delta function $\delta(t)$ of time t, the output is called "impulse response" of the operator, which we denote as $h(t)$:

$$h(t) = L[\delta(t)]. \tag{39}$$

Here, we introduce "convolution" $f(t) * g(t)$ of functions $f(t)$ and $g(t)$, which is defined by a following infinite integration:

$$f(t) * g(t) \equiv \int_{-\infty}^{\infty} f(t-\alpha)\, g(\alpha)\, d\alpha, \tag{40}$$

where symbol "$*$" stands for the operation of the convolution. Convolution has following properties:

$$f(t) * g(t) = g(t) * f(t) \qquad \text{(commutative)},$$

because

$$g(t) * f(t) = \int_{-\infty}^{\infty} g(t-\beta)\, f(\beta)\, d\beta = \int_{-\infty}^{\infty} f(\beta)\, g(t-\beta)\, d\beta = \int_{-\infty}^{\infty} f(t-\alpha)\, g(\alpha)\, d\alpha,$$

where we used a transformation of the argument of the integration: $\alpha = t - \beta$ and hence $d\beta = -d\alpha$,
and

$$f(t) * g(-t) = \int_{-\infty}^{\infty} f(t-\beta)\, g(-\beta)\, d\beta = \int_{-\infty}^{\infty} f(t+\alpha)\, g(\alpha)\, d\alpha,$$

where we used $\alpha = -\beta$ and $d\beta = -d\alpha$.

Then, the output of the linear system can be represented as a convolution of the input and the impulse response, i.e.,

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(t-\alpha)\, h(\alpha)\, d\alpha. \tag{41}$$

This equation is easily proven, based on the definition of the delta function, in the following way:

$$\begin{aligned} y(t) = L[x(t)] &= L[\int_{-\infty}^{\infty} x(\beta)\, \delta(t-\beta)\, d\beta] = \int_{-\infty}^{\infty} x(\beta)\, L[\delta(t-\beta)]\, d\beta \\ &= \int_{-\infty}^{\infty} x(\beta)\, h(t-\beta)\, d\beta = \int_{-\infty}^{\infty} x(t-\alpha)\, h(\alpha)\, d\alpha = x(t) * h(t). \end{aligned}$$

Note that L operates only on a function of time t.

Linear Systems with Random Processes as Inputs

Hereafter, we will consider linear systems having random processes as inputs. Then, we have following general properties.

- Expectation of the output:
  If $x(t)$ and $y(t)$ are input and output of a linear system, their expectations $\langle x(t) \rangle$ and $\langle y(t) \rangle$ are also related with each other as input and output of the same linear system, i.e.,

$$\begin{aligned} &\langle L[x(t)] \rangle = L[\langle x(t) \rangle], \\ &\text{or} \\ &\langle x(t) * h(t) \rangle = \langle x(t) \rangle * h(t), \\ &\text{or} \\ &\eta_y(t) = L[\eta_x(t)]. \end{aligned} \tag{42}$$

  Proof :

$$\begin{aligned} \langle x(t) * h(t) \rangle = \langle \int_{-\infty}^{\infty} x(t-\alpha)\, h(\alpha)\, d\alpha \rangle &= \int_{-\infty}^{\infty} \langle x(t-\alpha) \rangle\, h(\alpha)\, d\alpha \\ &= \langle x(t) \rangle * h(t). \end{aligned}$$

  Note that the impulse response $h(t)$ is a deterministic function of time, and, hence, not affected by the ensemble average.

- Autocorrelation of the output:

$$R_{yy}(t_1,\, t_2) = R_{xx}(t_1,\, t_2) * h(t_1) * h^*(t_2),$$
or
$$R_{yy}(t_1,\, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{xx}(t_1 - \alpha,\, t_2 - \beta)\, h(\alpha)\, h^*(\beta)\, d\alpha\, d\beta. \quad (43)$$

Proof :

1. Cross–correlation of the input and the output is

$$R_{xy}(t_1,\, t_2) = \langle x(t_1)\, y^*(t_2) \rangle = \langle x(t_1)\, x^*(t_2) * h^*(t_2) \rangle$$
$$= \langle x(t_1)\, x^*(t_2) \rangle * h^*(t_2) = R_{xx}(t_1,\, t_2) * h^*(t_2),$$
or
$$R_{xy}(t_1,\, t_2) = \int_{-\infty}^{\infty} R_{xx}(t_1,\, t_2 - \beta)\, h^*(\beta)\, d\beta.$$

2. Autocorrelation of the output is

$$R_{yy}(t_1,\, t_2) = \langle y(t_1)\, y^*(t_2) \rangle = \langle x(t_1) * h(t_1)\, y^*(t_2) \rangle$$
$$= \langle x(t_1)\, y^*(t_2) \rangle * h(t_1) = R_{xy}(t_1,\, t_2) * h(t_1),$$
or
$$R_{yy}(t_1,\, t_2) = \int_{-\infty}^{\infty} R_{xy}(t_1 - \alpha,\, t_2)\, h(\alpha)\, d\alpha.$$

3. From 1. and 2. above, we have

$$R_{yy}(t_1,\, t_2) = R_{xx}(t_1,\, t_2) * h^*(t_2) * h(t_1)$$
$$= R_{xx}(t_1,\, t_2) * h(t_1) * h^*(t_2),$$
or
$$R_{yy}(t_1,\, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{xx}(t_1 - \alpha,\, t_2 - \beta)\, h(\alpha)\, h^*(\beta)\, d\alpha\, d\beta.$$

1.2.3 Stationary Random Processes

Definitions

- A random process $z(t)$ is called "stationary" (or, more specifically, "wide–sense stationary"), if the expectation does not depend on time, and the autocorrelation is a function of time difference only:

$$\langle z(t) \rangle = \eta = \text{const},$$
$$\langle z(t_1)\, z^*(t_2) \rangle = R(\tau), \quad (44)$$

where $\tau \equiv t_1 - t_2$.

- Random processes $x(t)$ and $y(t)$ are called "jointly stationary", if both of them are stationary, and their cross–correlation is a function of time difference only:
$$\langle x(t_1)\, y^*(t_2)\rangle = R_{xy}(\tau), \tag{45}$$
where $\tau \equiv t_1 - t_2$.

Of course, it is not easy to find actual physical processes which strictly satisfy these conditions. For example, some of quasars or astronomical masers are known to exhibit significant time variations in yearly, or shorter, time scales. In a practical sense, however, physical processes are well approximated by the stationary random processes, if equations (44) and (45) are fulfilled during time scales, which are sufficient to estimate their statistical properties (see discussions on ergodicity in subsection 1.2.4). In this sense, many physical processes can be successfully modeled as stationary random processes.

Properties

Following formulae can be easily derived, by applying general properties of correlations, covariances, and so on, to the particular case of the stationary random processes as defined above.

- $R(-\tau) = R^*(\tau)$.
- $R(0) = \langle \mid z(t) \mid^2 \rangle \geq 0$.
- Positive definiteness:
$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i\, a_j^* R(t_i - t_j) \geq 0, \quad \text{for any } a_i.$$
- $\mid R(\tau) \mid \leq R(0)$.
- $C(\tau) = R(\tau) - \mid \eta \mid^2$, autocovariance.
- $r(\tau) = C(\tau)/C(0)$, correlation coefficient.
- $\mid r(\tau) \mid \leq 1$.
- $R_{xy}(-\tau) = R_{yx}^*(\tau)$.
- $\mid R_{xy}(\tau) \mid^2 \leq R_{xx}(0)\, R_{yy}(0)$.

- $C_{xy}(\tau) = R_{xy}(\tau) - \eta_x\,\eta_y^*$, cross–covariance.
- $r_{xy} = C_{xy}(\tau)/\sqrt{C_{xx}(0)\,C_{yy}(0)}$, cross–correlation coefficient.
- $|\ r_{xy}(\tau)\ | \leq 1$.

1.2.4 Ergodicity

How can we estimate various statistical properties of a random process, if we are given with a single time series only, which is an outcome of a single trial of the random process? As a matter of fact, you would better not to expect to get more than a day (i.e. a single trial) of telescope time for your VLBI observation of a certain object, unless referees of your proposal are extremely generous.

The answer to this question is, 'We can do it, if our random process is ergodic!'.

- Definition.
  A random process $z(t)$ is called "ergodic" if its ensemble averages are equal to appropriate time averages.

  > This implies that we can estimate any statistical property of $z(t)$, using time average of an outcome of a single trial, if the random process is ergodic.

- Mean–ergotic process.
  A random process $z(t)$, with a constant expectation:

  $$\eta = \langle z(t) \rangle,$$

  is called "mean–ergotic", if its time average η_T tends to η as averaging time $2T$ tends to infinity:

  $$\eta_T = \frac{1}{2T} \int_{-T}^{T} z(t)dt \ \ \rightarrow \ \ \eta \quad \text{as} \quad T \ \rightarrow \ \infty.$$

- A condition for the mean–ergotic process.
  It is evident that

  $$\langle \eta_T \rangle = \frac{1}{2T} \int_{-T}^{T} \langle z(t) \rangle dt = \eta.$$

Therefore, $z(t)$ is mean–ergotic, if the "variance", or "dispersion", σ_T^2 of the time average η_T tends to 0 as $T \to \infty$, i.e.,

$$\sigma_T^2 \equiv \langle \mid \eta_T - \eta \mid^2 \rangle \to 0 \ \text{ as } \ T \to \infty.$$

If we introduce the autocovariance $C(t_1,\, t_2)$ of $z(t)$, the above condition is equivalent to

$$\begin{aligned} \sigma_T^2 &= \frac{1}{(2T)^2} \int_{-T}^{T} \int_{-T}^{T} \langle (z(t_1) - \eta)\, (z(t_2) - \eta)^* \rangle \, dt_1 \, dt_2 \\ &= \frac{1}{(2T)^2} \int_{-T}^{T} \int_{-T}^{T} C(t_1,\, t_2) \, dt_1 \, dt_2 \to 0 \quad \text{ as } \ T \to \infty. \end{aligned}$$

- In the stationary random case.
  Strictly speaking, the ergodicity is inconceivable if $\langle z(t) \rangle$ varies in time. So, we assume that $z(t)$ is a stationary random process, and therfore, its autocovariance is a function of time diffference $\tau = t_1 - t_2$:

$$C(t_1,\, t_2) = C(\tau).$$

Then, the double integration in the above condition is reduced to a

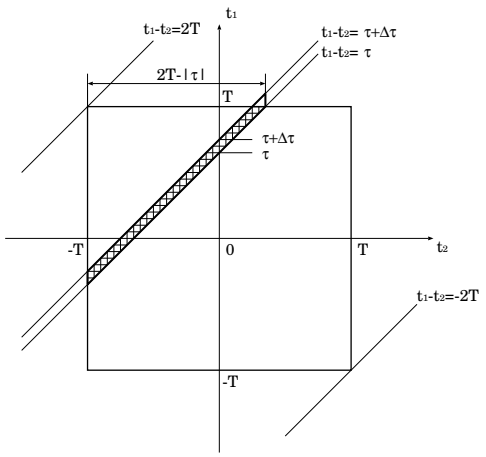

Figure 11: Geometry of the integration.

single one:

$$\frac{1}{2T} \int_{-2T}^{2T} C(\tau) \left(1 - \frac{\mid \tau \mid}{2T} \right) d\tau \to 0 \ \text{ as } \ T \to \infty, \tag{46}$$

because

$$\int_{-T}^{T}\int_{-T}^{T} C(t_1,\, t_2)\, dt_1\, dt_2 = \int_{-2T}^{2T} C(\tau)(2T- \mid \tau \mid)\, d\tau, \tag{47}$$

as we see from Figure 11. In fact, in the rectangular range of integration $-T \le t_1 \le T$ and $-T \le t_2 \le T$ in Figure 11, the autocovariance $C(\tau)$ is constant along a line $t_1 - t_2 = \tau$, and an area of the hatched region, sandwiched between two lines $t_1 - t_2 = \tau$ and $t_1 - t_2 = \tau + \Delta\tau$ is nearly equal to the area of the enclosing parallelogram, which is equal to $(2T- \mid \tau \mid)\, \Delta\tau$, in the linear approximation with respect to small $\Delta\tau$.

- Correlation–ergodic process.
  A stationary random process $z(t)$ with an autocorrelation

$$R(\xi) = \langle z(t+\xi)\, z^*(t) \rangle,$$

is called "correlation–ergodic", if a process defined by $u_\xi(t)$:

$$u_\xi(t) = z(t+\xi)\, z^*(t),$$

with an expectation $\langle u_\xi(t) \rangle = R(\xi)$, is mean–ergotic, i.e., if

$$R_T \equiv \frac{1}{2T} \int_{-T}^{T} u_\xi(t)\, dt \to R(\xi) \;\text{ as }\; T \to \infty.$$

Similarly to the case of the mean–ergotic process, $z(t)$ is correlation–ergotic, if the variance, or dispersion, $\sigma^2_{R_T} \equiv \langle \mid R_T - R(\xi) \mid^2 \rangle$ tends to 0 as $T \to \infty$, or, equivalently, if an autocovarinace of $u_\xi(t)$:

$$C_{uu}(\tau) = R_{uu}(\tau) - R^2(\xi),$$

where $R_{uu}(\tau)$ is an autocorrelation of $u_\xi(t)$:

$$R_{uu}(\tau) = \langle u_\xi(t+\tau)\, u^*_\xi(t) \rangle = \langle z(t+\xi+\tau)\, z^*(t+\tau)\, z^*(t+\xi)\, z(t) \rangle,$$

satisfies the condition in equation (46) for any ξ, i.e.,

$$\frac{1}{2T} \int_{-2T}^{2T} C_{uu}(\tau) \left(1 - \frac{\mid \tau \mid}{2T}\right) d\tau \to 0 \;\text{ as }\; T \to \infty. \tag{48}$$

- An illustration of the correlation–ergodic process.

  Let us try to do an exercise in order to simulate the correlation–ergodicity using a simple model of a normal (Gaussian) process.

  We generate a series of real zero–mean random numbers obeying the first–order Gaussian probability density with a dispersion σ^2, by means of the standard Box–Muller method (see, for example, Press et al., 1988). Then we regard the series as evenly ordered in time t within a range of data $0 < t \leq t_u$ with a certain interval ("sampling interval") Δt between two consecutive data points. Thus we obtain a model sample of a stationary normal process, which we call "sample data" and denote by $v_{rand}(t)$.

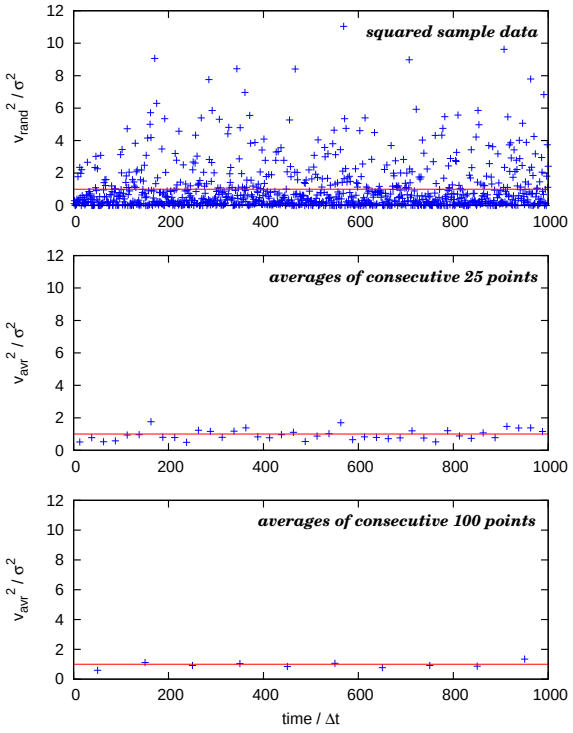

Figure 12: An illustration of the correlation–ergodicity. Top: time variation of squared sample data of a stationary normal process. Middle and Bottom: Time variations of averages of the squared sample data over 25 and 100 consecutive data points, respectively. All plotted values are normalized by the dispersion σ^2.

Figure 12 shows time variations of squared sample data $v^2_{rand}(t)$ (Top),

as well as time averages of the squared sample data:

$$v^2_{avr}(t,\,\tau_a) = \frac{1}{\tau_a}\int\limits_{t-\tau_a/2}^{t+\tau_a/2} v^2_{rand}(t')dt',$$

over 25 (Middle), and 100 (Bottom) consecutive data points, respectively, where τ_a is the averaging time, i.e. $\tau_a = N_{\tau_a}\,\Delta t$ if we denote the number of consecutive data points in the averaging time by N_{τ_a}.

It is evident from Figure 12 that the averaged squared sample data $v^2_{avr}(t,\,\tau_a)$ indeed converge to the dispersion of the original sample data $\langle v^2_{rand}(t)\rangle = \sigma^2$, or $v^2_{avr}\,/\,\sigma^2 \;\to\; 1$ in Figure 12, with increasing averaging time τ_a, as the correlation–ergodicity requires.

In order to see this convergence more clearly, we estimate the standard deviation $S_d(\tau_a) = \langle[v^2_{avr}(t,\,\tau_a) - \sigma^2]^2\rangle^{1/2}$ of the averaged squared sample data $v^2_{avr}(t,\,\tau_a)$ by calculating the root mean square $R_{ms}(\tau_a)$ of $v^2_{avr}(t,\,\tau_a) - \sigma^2$ over the whole data range $0 < t \le t_u$:

$$S_d(\tau_a) \cong R_{ms}(\tau_a) \equiv \left\{\frac{1}{t_u}\int\limits_0^{t_u}[v^2_{avr}(t,\,\tau_a) - \sigma^2]^2 dt\right\}^{1/2}.$$

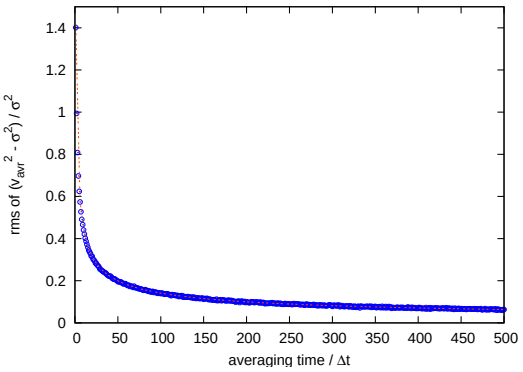

Figure 13: Root mean square of $v^2_{avr}(t,\,\tau_a) - \sigma^2$ normalised by the dispersion σ^2 of the original sample data: $\sigma^2 = \langle v^2_{rand}(t)\rangle$ (circles). This gives an estimation of the standard deviation of the time–averaged squared sample data $v^2_{avr}(t,\,\tau_a)$ which decreases with averaging time well following a curve $\sqrt{2}\,/\sqrt{N_{\tau_a}}$ (broken line), where N_{τ_a} is the number of data points in the averaging time, i.e. $N_{\tau_a} = \tau_a\,/\,\Delta t$.

Figure 13 shows the result of the calculation normalized by σ^2.

From Figure 13, we see that the standard deviation $S_d(\tau_a)$ of $v^2_{avr}(t,\, \tau_a)$ decreases with averaging time to zero as

$$S_d(\tau_a) = \frac{\sqrt{2}\,\sigma^2}{\sqrt{N_{\tau_a}}} = \frac{\sigma^2}{\sqrt{\frac{\tau_a}{2\Delta t}}},$$

just as required by the correlation–ergodicity and expected from the averaging law of independent noises. The standard deviation is equal to $\sqrt{2}\,\sigma^2$ at $N_{\tau_a} = 1$ in accordance with the expected one:

$$\left[\int_{-\infty}^{\infty} (x^2 - \sigma^2)^2 \,\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma^2}}\, dx\right]^{1/2} = \sqrt{2}\,\sigma^2.$$

Note that the ergodicity is guaranteed from the begining in the above exercise, since the "time series" we generated is actually an ensemble of normal random numbers, which are statistically independent samples with respect to each other. Therefore, the "time average" must naturally tend to the ensemble average when the averaging time, and the number of independent samples N_{τ_a}, too, increase. In this sense, the above exercise might be trivial and meaningless.

From another viewpoint, however, the exercise helps us to learn that there is nothing strange in the ergodicity if a sample time series forms an ensemble of all possible values of the random variable characterizing our random process. We can conceive that an ergodic process is a random process having such a property.

- Signal–to–noise ratio in mean power measurements.

  As we noticed earlier, the radio waves from astronomical radio sources are random noise time series which are hardly distinguishable from receiver and other noises by their temporal behaviors. Strength of such a "noise signal" is given by its mean power, i.e. the dispersion. In order to estimate the mean power, we measure a "time–averaged power", i.e. a time average of the squared noise signal, as we discussed in Chapter 2. In the context of present discussions, we assume the ergodicity of the noise signal and approximate an ensemble average by a time average. As a simple example of the ergodic process, we adopt here the normal process model which we used in the above exercise.

  As far as the averaging time τ_a is finite, the time–averaged power of

the noise signal $s(t)$:

$$s^2_{avr}(t,\tau_a) = \frac{1}{\tau_a}\int\limits_{t-\tau_a/2}^{t+\tau_a/2} s^2(t')dt',$$

fluctuates around its mean, i.e. the dispersion $\sigma_s^2 = \langle s^2(t)\rangle$. Strength of such a fluctuation is characterized by a standard deviation $S_{d_s}(\tau_a) = \langle[s^2_{avr}(t,\,\tau_a) - \sigma_s^2]^2\rangle^{1/2}$ which is reduced to zero with increasing τ_a as

$$S_{d_s}(\tau_a) = \frac{\sqrt{2}\,\sigma_s^2}{\sqrt{N_{\tau_a}}} = \frac{\sigma_s^2}{\sqrt{\frac{\tau_a}{2\Delta t}}},$$

where Δt and N_{τ_a} are again the "sampling interval" and the number of independent samples within the averaging time, respectively.

For the mean power measurement, we can treat the dispersion σ_s^2 as a "signal", and the standard deviation of the time–averaged power as a "noise".

Therefore, if we had only the noise signal $s(t)$ from a radio source, and did not have any system noise, the signal–to–noise ratio would be given by:

$$(S/N)_s = \frac{\sigma_s^2}{S_{d_s}(\tau_a)} = \sqrt{\frac{N_{\tau_a}}{2}} = \sqrt{\frac{\tau_a}{2\Delta t}}.$$

This would imply that we could get a fairly high signal–to–noise ratio $(S/N)_s = 5$ after a time–averaging over $N_{\tau_a} = 50$ independent samples in this "signal only" case.

However, with actual radio telescopes, we always observe superposed noisy signal $s(t)$ and system noise $n(t)$, which can be regarded as two independent ergodic processes. Usually, the system noise occuring mainly in receivers and atmosphere, is much stronger than the noisy signal from a radio source, as we saw in Chapter 2. This means that the "noise dispersion" $\sigma_n^2 = \langle n^2(t)\rangle$ is much larger than the "signal dispersion" σ_s^2.

When we measure the mean power of the signal plus noise data, as we do in the single–dish radiometry, the "noise" in such a measurement is mostly determined by the fluctuation of the time–averaged power of the system noise:

$$n^2_{avr}(t,\tau_a) = \frac{1}{\tau_a}\int\limits_{t-\tau_a/2}^{t+\tau_a/2} n^2(t')dt'.$$

In our normal process model, the standard deviation of the fluctuating $n^2_{avr}(t, \tau_a)$ is given by

$$S_{d_n}(\tau_a) = \frac{\sqrt{2}\,\sigma_n^2}{\sqrt{N_{\tau_a}}} = \frac{\sigma_n^2}{\sqrt{\frac{\tau_a}{2\Delta t}}},$$

while the "signal", which we detect by the "on–off" observation, is still given by the dispersion of the radio source signal σ_s^2. Therefore, the signal–to–noise ratio is now given by

$$S/N = \frac{\sigma_s^2}{S_{d_n}(\tau_a)} = \frac{\sigma_s^2}{\sigma_n^2}\sqrt{\frac{N_{\tau_a}}{2}} = \frac{\sigma_s^2}{\sigma_n^2}\sqrt{\frac{\tau_a}{2\Delta t}}.$$

The mean powers of the signal and the noise are described in terms of the antenna temperature T_A and the system noise temperature T_S, respectively, as we saw in Chapter 2. When we observe a continuum source with a telescope having a frequency bandwidth B, we have

$$\sigma_s^2 = kT_AB, \quad \text{and} \quad \sigma_n^2 = kT_SB,$$

where $k = 1.381 \times 10^{-23}$ J K^{-1} is the Boltzmann constant as before. Then, the signal–to–noise ratio is given by

$$S/N = \frac{T_A}{T_S}\sqrt{\frac{N_{\tau_a}}{2}} = \frac{T_A}{T_S}\sqrt{\frac{\tau_a}{2\Delta t}}.$$

Thus, owing to the ergodicity, we can measure the mean power of the radio source signal, improving the signal–to–noise ratio in proportion to the square root of the averaging time, even when the signal is buried in much higher system noise. However, the averaging time required for achieving a sufficiently high signal–to–noise ratio is much longer than the "signal only" case, by a factor of $(T_S/T_A)^2$. For example, if $T_S/T_A = 100$, we need $N_{\tau_a} = 500,000$ independent samples, instead of 50 in the "signal only" case, to get $S/N = 5$. This "additional integration" necessary for suppressing the system noise effect is an essential part of data processing in radio astronomy.

Note that the last equation is consistent with that for the signal–to–noise ratio in the single–dish radiometry, which we saw in Chapter 2.

- Radio astronomical data and ergodicity.

  For most of actual physical processes, it is likely that their autocovariances are finite everywhere and tend to 0 when the time difference

argument τ tends to infinity. Therefore, equations (46) and (48), which are conditions for establishment of the ergodicity, appear well satisfied in the most cases. Moreover, it is well known that the signal–to–noise ratio of the "power" or the "correlation" of a moderately strong signal from an astronomical radio source, obtained in a square–law detector or in a correlator, improves with averaging (or integration) time τ_a nearly following the $\propto \sqrt{\tau_a}$ law. This means that the simple normal process model mentioned above is fairly a good model of actually observed data of radio astronomy. On the other hand, the "power", or "correlation", thus obtained is usually almost time–invariant during time–scales from hours to months. Therefore, radio astronomical data are mostly consistent with assumptions of the stationary random process and the ergodicity.

1.2.5 Stationary Random Processes in Linear Systems

Let us consider cases when inputs of linear systems are stationary random processes. Then, we have following properties.

- If an input $x(t)$ in a linear system $y(t) = L[x(t)]$ with an impulse response $h(t)$ is a stationary random process, then an output $y(t)$ is also a stationary random process.
  Proof :

  1. Expectation $\eta_y(t)$ of the output $y(t)$ is constant in time. In fact,

  $$\eta_y(t) = \int_{-\infty}^{\infty} \eta_x(t-\alpha)\, h(\alpha)\, d\alpha = \eta_x \int_{-\infty}^{\infty} h(\alpha)\, d\alpha = \text{const}, \qquad (49)$$

  since $x(t)$ is stationary and therefore $\eta_x(t) = \eta_x$ is constant in time. So, we have $\eta_y(t) = \eta_y = \text{const}$.

  2. Autocorrelation $R_{yy}(t_1,\, t_2)$ of the output $y(t)$ is a function of time difference $\tau = t_1 - t_2$ only, because

  $$\begin{aligned} R_{yy}(t_1,\, t_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{xx}(t_1-\alpha,\, t_2-\beta)\, h(\alpha)\, h^*(\beta)\, d\alpha\, d\beta \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{xx}(\tau-\alpha+\beta)\, h(\alpha)\, h^*(\beta)\, d\alpha\, d\beta, \end{aligned}$$

  since $x(t)$ is stationary and therefore

  $$R_{xx}(t_1-\alpha,\, t_2-\beta) = R_{xx}(t_1-\alpha-(t_2-\beta)) = R_{xx}(\tau-\alpha+\beta).$$

Then, the above formula is now expressed as:

$$R_{yy}(\tau) = R_{xx}(\tau) * h(\tau) * h^*(-\tau). \tag{50}$$

- If an input $x(t)$ in a linear system $y(t) = L[x(t)]$ is a stationary random process, then the input $x(t)$ and the output $y(t)$ are jointly stationary.

  Proof :

  1. We have proven above that, if the input is stationary, then the output is also stationary, i.e., both $x(t)$ and $y(t)$ are stationary.
  2. Cross–correlation $R_{xy}(t_1,\, t_2)$ of the input $x(t)$ and the output $y(t)$ is a function of time difference $\tau = t_1 - t_2$ only, because

$$\begin{aligned} R_{xy}(t_1,\, t_2) = \langle x(t_1)\, y^*(t_2)\rangle &= \int_{-\infty}^{\infty} \langle x(t_1)\, x^*(t_2-\alpha)\rangle\, h^*(\alpha)\, d\alpha \\ &= \int_{-\infty}^{\infty} R_{xx}(\tau+\alpha)\, h^*(\alpha)\, d\alpha = R_{xx}(\tau) * h^*(-\tau), \end{aligned}$$

  where $h(t)$ is an impulse response of L as before. Thus, we have

$$R_{xy}(\tau) = R_{xx}(\tau) * h^*(-\tau). \tag{51}$$

- Likewise, we can prove an equation:

$$R_{yy}(\tau) = R_{xy}(\tau) * h(\tau), \tag{52}$$

  which, together with equation (51), offers another derivation of equation (50).

- If $x_1(t)$ and $x_2(t)$ are jointly stationary random processes, then outputs $y_1(t)$ and $y_2(t)$, which are obtained from $x_1(t)$ and $x_2(t)$ through arbitrary linear operators L_1 and L_2 with impulse responses $h_1(t)$ and $h_2(t)$, respectively, are also jointly stationary.

  Proof :

  We have

$$y_1(t) = L_1[x_1(t)] = x_1(t) * h_1(t) = \int_{-\infty}^{\infty} x_1(t-\alpha)\, h_1(\alpha)\, d\alpha,$$

$$y_2(t) = L_2[x_2(t)] = x_2(t) * h_2(t) = \int_{-\infty}^{\infty} x_2(t-\alpha)\, h_2(\alpha)\, d\alpha.$$

Both $y_1(t)$ and $y_2(t)$ are, of course, stationary, and their cross–correlation:

$$\begin{aligned} R_{y_1y_2}(t_1,\, t_2) &= \langle y_1(t_1)\, y_2^*(t_2)\rangle \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \langle x_1(t_1-\alpha)\, x_2^*(t_2-\beta)\rangle\, h_1(\alpha)\, h_2^*(\beta)\, d\alpha\, d\beta \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{x_1x_2}(\tau-\alpha+\beta)\, h_1(\alpha)\, h_2^*(\beta)\, d\alpha\, d\beta, \end{aligned}$$

is a function of time difference $\tau = t_1 - t_2$ only. This proves the joint stationarity of $y_1(t)$ and $y_2(t)$, and yields

$$R_{y_1y_2}(\tau) = R_{x_1x_2}(\tau) * h_1(\tau) * h_2^*(-\tau). \tag{53}$$

1.2.6 Spectra of Stationary Random Processes

Definitions

- Fourier transform $S(\omega)$ of an autocorrelation $R(\tau)$ of a stationary random process is called "power spectrum" (or "spectral density") of the process. Here, ω is an angular frequency, which is related to a linear frequency ν as $\omega = 2\pi\nu$. Thus, the power spectrum and the autocorrelation are related to each other by the Fourier– and inverse Fourier transforms:

$$S(\omega) = \int_{-\infty}^{\infty} R(\tau)e^{-i\omega\tau}\, d\tau, \tag{54}$$

$$R(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)e^{i\omega\tau}\, d\omega. \tag{55}$$

  Hereafter, we express a Fourier transform pair by a symbol "⇔". Then,

$$S(\omega) \Leftrightarrow R(\tau).$$

- Fourier transform $S_{xy}(\omega)$ of a cross–correlation $R_{xy}(\tau)$ of jointly stationary random processes $x(t)$ and $y(t)$ is called "cross–power spectrum". Thus,

$$S_{xy}(\omega) = \int_{-\infty}^{\infty} R_{xy}(\tau)e^{-i\omega\tau}\, d\tau, \tag{56}$$

$$R_{xy}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xy}(\omega)e^{i\omega\tau}\, d\omega, \tag{57}$$

and

$$S_{xy}(\omega) \Leftrightarrow R_{xy}(\tau).$$

Note that convergence of Fourier integrals in equations (54) and (56), and therefore in their inverses in equations (55) and (57), too, is usually guaranteed, since, for actual physical processes, $R(\tau)$ and $R_{xy}(\tau)$ are mostly finite everywhere, and tend to zero as $\tau \to \pm\infty$.

Properties

- **Power** $\langle |\ z(t)\ |^2 \rangle$ of a stationary random process $z(t)$ is equal to an integrated power spectrum over the whole frequency range:

$$\langle |\ z(t)\ |^2 \rangle = R(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega)\, d\omega = \int_{-\infty}^{\infty} S(2\pi\nu)\, d\nu, \qquad (58)$$

  where $\nu = \omega/(2\pi)$ is a frequency, corresponding to the angular frequency ω. Thus, power spectrum is really a "spectrum of the power".

- Power spectrum $S(\omega)$ is a real function.

  Proof :

  Since $R(-\tau) = R^*(\tau)$, complex conjugate of $S(\omega)$ is equal to itself because

$$S^*(\omega) = \int_{-\infty}^{\infty} R^*(\tau) e^{i\omega\tau}\, d\tau = \int_{-\infty}^{\infty} R(-\tau) e^{i\omega\tau}\, d\tau = \int_{-\infty}^{\infty} R(\tau) e^{-i\omega\tau}\, d\tau = S(\omega).$$

- For any cross–power spectrum, $S_{xy}(\omega) = S^*_{yx}(\omega)$.

  Proof :

  Since $R_{xy}(-\tau) = R^*_{yx}(\tau)$,

$$S^*_{yx}(\omega) = \int_{-\infty}^{\infty} R^*_{yx}(\tau) e^{i\omega\tau}\, d\tau = \int_{-\infty}^{\infty} R_{xy}(-\tau) e^{i\omega\tau}\, d\tau = \int_{-\infty}^{\infty} R_{xy}(\tau) e^{-i\omega\tau}\, d\tau$$
$$= S_{xy}(\omega).$$

- A power spectrum $S(\omega)$ corresponding to a real autocorrelation $R(\tau)$ is an even function of ω (see Figure 14):

$$S(-\omega) = S(\omega). \qquad (59)$$

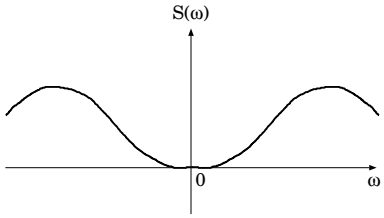

Figure 14: Power spectrum is even when autocorrelation is real.

Proof :

Since, in this case, $R(-\tau) = R(\tau)$ (the real autocorrelation is an even function of τ),

$$S(-\omega) = \int_{-\infty}^{\infty} R(\tau) e^{i\omega\tau}\, d\tau = \int_{-\infty}^{\infty} R(-\tau) e^{i\omega\tau}\, d\tau = \int_{-\infty}^{\infty} R(\tau) e^{-i\omega\tau}\, d\tau = S(\omega).$$

- A cross–power spectrum corresponding to a real cross–correlation satisfies

$$S_{xy}(-\omega) = S_{yx}(\omega), \tag{60}$$

and, therefore, is Hermitian symmetric:

$$S_{xy}(-\omega) = S^*_{xy}(\omega), \tag{61}$$

(see Figure 15).

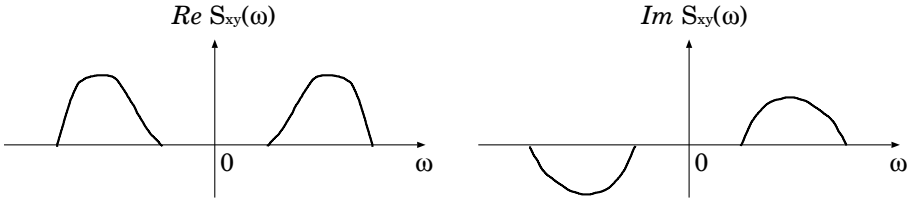

Figure 15: Cross–power spectrum is Hermitian symmetric (i.e., real part is even and imaginary part is odd) when cross–correlation is real.

Proof :

Since, in this case, $R_{xy}(-\tau) = R_{yx}(\tau)$,

$$S_{xy}(-\omega) = \int_{-\infty}^{\infty} R_{xy}(\tau) e^{i\omega\tau}\, d\tau = \int_{-\infty}^{\infty} R_{xy}(-\tau) e^{-i\omega\tau}\, d\tau = \int_{-\infty}^{\infty} R_{yx}(\tau) e^{-i\omega\tau}\, d\tau$$
$$= S_{yx}(\omega),$$

and, in view of the general property $S_{xy}(\omega) = S^*_{yx}(\omega)$, we also have

$$S_{xy}(-\omega) = S^*_{xy}(\omega).$$

- Real autocorrelation can be described solely by the positive frequency range of the power spectrum.

  Proof :

  Since $S(\omega) = S(-\omega)$ in this case,

$$\begin{aligned} R(\tau) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)e^{i\omega\tau}\,d\omega = \frac{1}{2\pi}\left[\int_{-\infty}^{0} S(\omega)e^{i\omega\tau}\,d\omega + \int_{0}^{\infty} S(\omega)e^{i\omega\tau}\,d\omega\right] \\ &= \frac{1}{2\pi}\int_{0}^{\infty}[S(-w)e^{-i\omega\tau} + S(\omega)e^{i\omega\tau}]\,d\omega = \frac{1}{2\pi}\int_{0}^{\infty} S(\omega)[e^{-i\omega\tau} + e^{i\omega\tau}]\,d\omega \\ &= \frac{1}{\pi}\int_{0}^{\infty} S(\omega)\cos(\omega\tau)\,d\omega = \frac{1}{\pi}\Re\left[\int_{0}^{\infty} S(\omega)e^{i\omega\tau}\,d\omega\right]. \end{aligned} \tag{62}$$

- Real cross–correlation can be described solely by the positive frequency range of the cross–power spectrum.

  Proof :

  Since $S_{xy}(-\omega) = S^*_{xy}(\omega)$ in this case,

$$\begin{aligned} R_{xy}(\tau) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xy}(\omega)e^{i\omega\tau}\,d\omega = \frac{1}{2\pi}\left[\int_{-\infty}^{0} S_{xy}(\omega)e^{i\omega\tau}\,d\omega + \int_{0}^{\infty} S_{xy}(\omega)e^{i\omega\tau}\,d\omega\right] \\ &= \frac{1}{2\pi}\int_{0}^{\infty}[S_{xy}(-w)e^{-i\omega\tau} + S_{xy}(\omega)e^{i\omega\tau}]\,d\omega \\ &= \frac{1}{2\pi}\int_{0}^{\infty}[S^*_{xy}(\omega)e^{-i\omega\tau} + S_{xy}(\omega)e^{i\omega\tau}]\,d\omega \\ &= \frac{1}{\pi}\Re\int_{0}^{\infty} S_{xy}(\omega)e^{i\omega\tau}\,d\omega. \end{aligned} \tag{63}$$

- **White noise**: if the spectrum is flat throughout the whole frequency range, then the correlation is proportional to the delta function of τ.

  If $S(\omega) = S = \text{const}$, then

$$R(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)e^{i\omega\tau}\,d\omega = S\,\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega\tau}\,d\omega = S\,\delta(\tau). \tag{64}$$

If $S_{xy}(\omega) = S_{xy} = \text{const}$, then

$$R_{xy}(\tau) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} S_{xy}(\omega) e^{i\omega\tau}\, d\omega = S_{xy}\, \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{i\omega\tau}\, d\omega = S_{xy}\, \delta(\tau). \quad (65)$$

Here we used a formula

$$\int\limits_{-\infty}^{\infty} e^{i\omega\tau}\, d\omega = 2\pi\delta(\tau), \quad (66)$$

which is known as one of the definitions of Dirac's delta function.

Thus, if spectra of jointly stationary random processes are completely flat (white), then their correlations are non–zero, only when $\tau = 0$. The "geodetic explanation of VLBI" uses this property as we saw before.

- **Convolution theorem**: Fourier transform of a convolution of two functions is equal to a product of Fourier transforms of those functions, i.e., if $a(\tau) \Leftrightarrow A(\omega)$ and $b(\tau) \Leftrightarrow B(\omega)$, then

$$a(\tau) * b(\tau) \Leftrightarrow A(\omega)\, B(\omega). \quad (67)$$

Proof :

$$\int\limits_{-\infty}^{\infty} a(\tau) * b(\tau)\, e^{-i\omega\tau}\, d\tau = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} a(\tau - \alpha)\, b(\alpha)\, e^{-i\omega\tau}\, d\alpha\, d\tau$$

$$= \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} a(\tau')\, e^{-i\omega\tau'}\, b(\tau'')\, e^{-i\omega\tau''}\, d\tau'\, d\tau'' = A(\omega)\, B(\omega),$$

where we introduced transformations $\tau' = \tau - \alpha$ and $\tau'' = \alpha$.

- Another convolution theorem holds for a product of functions $a(\tau)$ and $b(\tau)$:

$$a(\tau)\, b(\tau) \Leftrightarrow \frac{1}{2\pi}\, A(\omega) * B(\omega), \quad (68)$$

because

$$\int\limits_{-\infty}^{\infty} a(\tau)\, b(\tau)\, e^{-i\omega\tau}\, d\tau$$

$$= \frac{1}{(2\pi)^2} \int\limits_{-\infty}^{\infty} \left[\int\limits_{-\infty}^{\infty} A(\omega')\, e^{i\omega'\tau}\, d\omega' \right] \left[\int\limits_{-\infty}^{\infty} B(\omega'')\, e^{i\omega''\tau}\, d\omega'' \right] e^{-i\omega\tau}\, d\tau$$

$$= \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(\omega')\, B(\omega'') \left[\int_{-\infty}^{\infty} e^{-i(\omega-\omega'-\omega'')\,\tau}\, d\tau \right] d\omega'\, d\omega''$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(\omega')\, B(\omega'')\, \delta(\omega-\omega'-\omega'')\, d\omega'\, d\omega''$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} A(\omega-\omega')\, B(\omega')\, d\omega' = \frac{1}{2\pi} A(\omega) * B(\omega),$$

where we used the relation

$$\int_{-\infty}^{\infty} e^{-i\omega\tau}\, d\tau = \int_{-\infty}^{\infty} e^{i\omega\tau}\, d\tau = 2\pi\delta(\omega).$$

- **Shift theorem**: If $a(\tau) \Leftrightarrow A(\omega)$, then

$$\begin{aligned} a(\tau-\tau_0) &\Leftrightarrow A(\omega)\, e^{-i\omega\tau_0}, \\ a(\tau) e^{i\omega_0\tau} &\Leftrightarrow A(\omega-\omega_0). \end{aligned} \tag{69}$$

Proof :

$$\int_{-\infty}^{\infty} a(\tau-\tau_0)\, e^{-i\omega\tau}\, d\tau = \int_{-\infty}^{\infty} a(\tau')\, e^{-i\omega(\tau'+\tau_0)}\, d\tau'$$

$$= \left[\int_{-\infty}^{\infty} a(\tau') e^{-i\omega\tau'}\, d\tau' \right] e^{-i\omega\tau_0} = A(\omega) e^{-i\omega\tau_0},$$

and,

$$\int_{-\infty}^{\infty} a(\tau) e^{i\omega_0\tau}\, e^{-i\omega\tau}\, d\tau = \int_{-\infty}^{\infty} a(\tau) e^{-i(\omega-\omega_0)\tau}\, d\tau = A(\omega-\omega_0). \tag{70}$$

1.2.7 Spectra of Outputs of Linear Systems

System Function

Let us call a Fourier transform $H(\omega)$ of an impulse response $h(t)$ of a linear system as the "system function":

$$H(\omega) = \int_{-\infty}^{\infty} h(t)\, e^{-i\omega t}\, dt, \tag{71}$$

or

$$H(\omega) \Leftrightarrow h(t).$$

For the system function, we have a general property common to all Fourier transforms:

$$H^*(\omega) \Leftrightarrow h^*(-t),$$

because

$$H^*(\omega) = \int_{-\infty}^{\infty} h^*(t) e^{i\omega t}\, dt = \int_{-\infty}^{\infty} h^*(-t) e^{-i\omega t}\, dt.$$

Furthermore, if the impulse response $h(t)$ is real, then

$$H^*(\omega) = \int_{-\infty}^{\infty} h(t) e^{i\omega t}\, dt = H(-\omega), \tag{72}$$

and, therefore, $\mid H(\omega) \mid^2$ is an even function of ω, because

$$\mid H(-\omega) \mid^2 = H(-\omega)\, H^*(-\omega) = H^*(\omega)\, H(\omega) = \mid H(\omega) \mid^2 . \tag{73}$$

Now, let us consider a linear system $y(t) = L[x(t)]$ with an impulse response $h(t)$ and a corresponding system function $H(\omega)$ with a stationary random process $x(t)$ as an input.

- Expectation of the output.

$$\eta_y = \eta_x \int_{-\infty}^{\infty} h(\alpha)\, d\alpha = \eta_x\, H(0). \tag{74}$$

- Power spectra of the output $S_{yy}(\omega)$ and the inputs $S_{xx}(\omega)$ are related to each other as

$$S_{yy}(\omega) = S_{xx}(\omega) \mid H(\omega) \mid^2 . \tag{75}$$

Proof :

In view of the convolution theorem given in equation (67), and properties of correlations,

$$\begin{aligned} R_{xy}(\tau) = R_{xx}(\tau) * h^*(-\tau) \quad &\Leftrightarrow \quad S_{xy}(\omega) = S_{xx}(\omega)\, H^*(\omega), \\ R_{yy}(\tau) = R_{xy}(\tau) * h(\tau) \quad &\Leftrightarrow \quad S_{yy}(\omega) = S_{xy}(\omega)\, H(\omega), \end{aligned}$$

and, hence

$$R_{yy}(\tau) = R_{xx}(\tau) * h(\tau) * h^*(-\tau) \quad \Leftrightarrow \quad S_{yy}(\omega) = S_{xx}(\omega) \mid H(\omega) \mid^2 .$$

- Autocorrelation of the outputs.

$$R_{yy}(\tau) = \langle y(t+\tau)\, y^*(t) \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{xx}(\omega) \mid H(\omega) \mid^2 e^{i\omega\tau}\, d\omega,$$

and, in particular,

$$R_{yy}(0) = \langle \mid y(t) \mid^2 \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{xx}(\omega) \mid H(\omega) \mid^2 d\omega. \qquad (76)$$

- Cross–power spectrum of outputs $y_1(t) = x_1(t) * h_1(t)$ and $y_2(t) = x_2(t) * h_2(t)$ of jointly stationary inputs $x_1(t)$ and $x_2(t)$ through two linear systems with impulse responses $h_1(t)$ and $h_2(t)$.

  As we saw earlier in equation (53), the cross–correlation $R_{y_1y_2}(\tau)$ of the outputs is expressed through the cross–correlation of the inputs $R_{x_1x_2}(\tau)$ as

$$R_{y_1y_2}(\tau) = R_{x_1x_2}(\tau) * h_1(\tau) * h_2^*(-\tau).$$

  Therefore, the convolution theorem in equation (67) gives us the cross–power spectrum:

$$S_{y_1y_2}(\omega) = S_{x_1x_2}(\omega) H_1(\omega) H_2^*(\omega), \qquad (77)$$

  where $S_{x_1x_2}(\omega) \Leftrightarrow R_{x_1x_2}(\tau)$ is a cross–power spectrum of the inputs.

- Cross–correlation of the outputs.

$$R_{y_1y_2}(\tau) = \langle y_1(t+\tau)\, y_2^*(t) \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{x_1x_2}(\omega) H_1(\omega)\, H_2^*(\omega) e^{i\omega\tau}\, d\omega,$$

  and, in particular,

$$R_{y_1y_2}(0) = \langle y_1(t)\, y_2^*(t) \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{x_1x_2}(\omega) H_1(\omega)\, H_2^*(\omega)\, d\omega. \qquad (78)$$

1.2.8 Two Designs of Spectrometers

As an example of applications of the theory of stationary random process, let us consider principles of two types of spectrometers which have been widely used in the radio astronomy (Figures 16 and 17).

In the **filterbank spectrometer** (Figure 16), received voltage from a radio source is equally fed to n identical analog narrow–band BPF's (band–pass–filters), which are called "filterbank" with successive center frequences

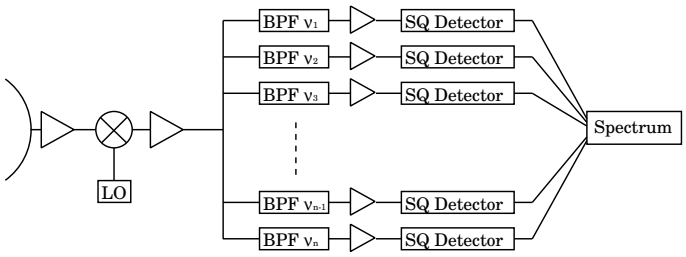

Figure 16: Basic design of a filterbank spectrometer.

$\nu_1, \nu_2, \cdots, \nu_n$. Outputs of the BPF's are squared and averaged by SQ (square–law) detectors and resultant powers yield a spectral shape of the source at the above frequencies.

In the **autocorrelation spectrometer** (Figure 17), on the other hand, the received voltage is first digitized by an analog–to–digital converter (A/D), and then equally divided into two digital signals, which are fed to n multipliers and integrators, one directly, and another with successive time delays $0, \tau, 2\tau, \cdots, (n-1)\tau$. The resultant 'autocorrelation' as a function of time delay is then Fourier transformed, and converted to a power spectrum.

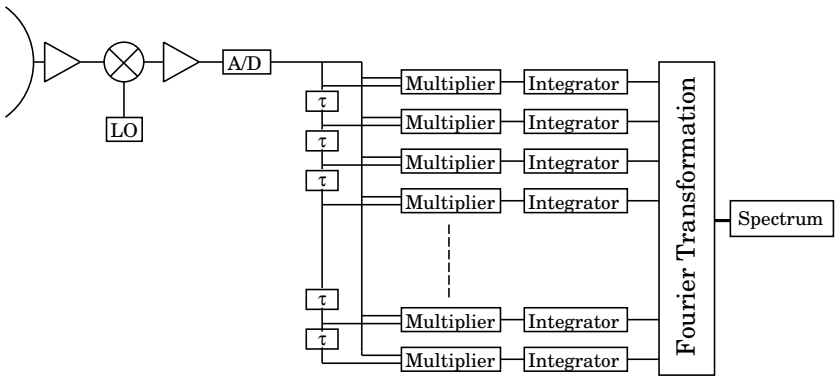

Figure 17: Basic design of an autocorrelation spectrometer.

The principles of the two designs look quite different. Do they really produce the same spectrum?

As far as the ergodicity holds, it is clear that the autocorrelation spectrometer must closely approximate the calculation of the power spectrum of the input signal (received voltage) as the Fourier transform of the autocorrelation, as we have described so far.

For the filterbank spectrometer, let us consider i-th narrow–band BPF as

a linear system, which has an input stationary random process $x(t)$, which is the received voltage in this case, an output $y_i(t)$, and an impulse response $h_i(t)$, having a rectangular sytem function $H_i(\omega)$:

$$H_i(\omega) = \begin{cases} \sqrt{\frac{2\pi}{\Delta\omega}} & \omega_i - \frac{\Delta\omega}{2} \leq \omega \leq \omega_i + \frac{\Delta\omega}{2}, \\ 0 & \text{otherwise}, \end{cases}$$

where $\omega_i = 2\pi\nu_i$ is the i-th center angular frequency, and $\Delta\omega$ is the frequency bandwidth of the BPF. If the power spectrum of the input is $S_{xx}(\omega)$, then, according to equation (75), the power spectrum of the output $S_{yy}(\omega)$ is

$$S_{yy}(\omega) = S_{xx}(\omega) \mid H_i(\omega) \mid^2,$$

(Figure 18). Since, in view of the ergodicity, the time averaging in a square–

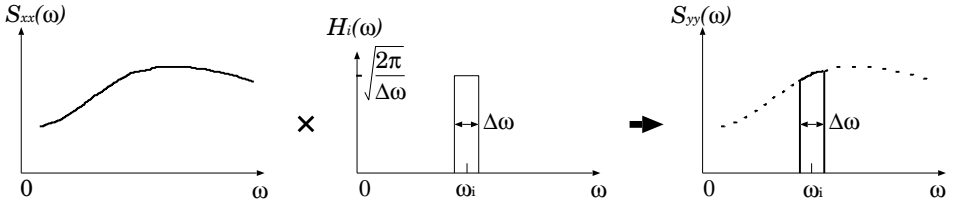

Figure 18: Band–pass filter passes a segment of the input power spectrum.

law detector must yield the power, or the autocorrelation at $\tau = 0$, of the output signal, if the averaging time is sufficiently long, we obtain

$$\langle \mid y_i(t) \mid^2 \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{yy}(\omega)\, d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{xx}(\omega) \mid H_i(\omega) \mid^2 d\omega$$

$$= \frac{1}{\Delta\omega} \int_{\omega_i - \frac{\Delta\omega}{2}}^{\omega_i + \frac{\Delta\omega}{2}} S_{xx}(\omega)\, d\omega.$$

This "power passed by a BPF" is nothing but a mean of the power spectrum of the received voltage $S_{xx}(\omega)$, involved in the spectral range $\omega_i - \frac{\Delta\omega}{2} \leq \omega \leq \omega_i + \frac{\Delta\omega}{2}$. Therefore, if $\Delta\omega$ is sufficiently narrow, and $S_{xx}(\omega)$ is continuous around ω_i, then we approximately have

$$\langle \mid y_i(t) \mid^2 \rangle \simeq S_{xx}(\omega_i).$$

Thus two spectrometers really yield the same power spectrum of the received voltage.

This example gives us a clear image that the power spectrum, defined as a Fourier transform of the autocorrelation of the input signal, is really a "spectrum of the power" of the signal.

1.2.9 Fourier Transforms of Stationary Random Processes

So far, we have considered Fourier transformation of correlations of the stationary random processes. Now, let us proceed to consider the Fourier transformation of the stationary random processes themselves.

Assume that a Fourier integral of a random process $z(t)$ is given by

$$Z(\omega) = \int_{-\infty}^{\infty} z(t)\, e^{-i\omega t}\, dt. \tag{79}$$

Since $z(t)$ is a random process in time t, it is natural to consider that $Z(\omega)$ is a random process in angular frequency ω, i.e., it is a function of ω, and its value at any ω is a random variable, which may vary from trial to trial.

If we apply the inverse Fourier transform to equation (79), we would have

$$z(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Z(\omega)\, e^{i\omega t}\, d\omega, \tag{80}$$

i.e., we could express any random process in time t as a superposition of infinite number of frequency components, which are themselves random processes in angular frequency ω.

Strictly speaking, however, we must be aware that the convergence of the intrgrals in equations (79) and (80) is not, in general, guaranteed, since the random processes may have finite amplitudes from the infinite past to the infinite future.

Of course, we could restrict the actual integration range to $-T < t \leq T$ with sufficiently large T in order to force the integrals to converge. In fact, durations of actual physical processes are most likely to be shorter than the age of our Universe. However, a too strong emphasis on this point may cause difficulties when we require stationarity to the random processes.

Special integral forms are often introduced in the literature to assure the convergence. We will, however, just assume some kind of convergence of the above integrals, without being heavily involved in the mathematical strictness. Instead, we will concentrate our attentions to several simple but useful statistical relations between the random process $z(t)$ and its Fourier transform $Z(\omega)$.

Properties of Fourier transforms of the random processes.

- Expectation of $Z(\omega)$ is a Fourier transform of the expectation $\eta(t)$ of $z(t)$.

  Proof :

  Taking ensemble average of the two sides of the Fourier transformation in equation (79), we have

$$\langle Z(\omega)\rangle = \int_{-\infty}^{\infty} \langle z(t)\rangle\, e^{-i\omega t}\, dt = \int_{-\infty}^{\infty} \eta(t)\, e^{-i\omega t}\, dt.$$

- If $z(t)$ is a stationary random process, the expectation of $Z(\omega)$ has a delta–function form with respect to ω.

  Proof :

  Since

$$\langle z(t)\rangle = \eta = \text{const},$$

  we have

$$\langle Z(\omega)\rangle = \int_{-\infty}^{\infty} \eta\, e^{-i\omega t}\, dt = 2\pi\, \eta\, \delta(\omega), \tag{81}$$

  according to equation (66).

- An autocorrelation of $Z(\omega)$, defined as $\langle Z(\omega_1)\, Z^*(\omega_2)\rangle$, is related to a two–dimensional Fourier transform $\Gamma(\omega_1,\, \omega_2)$ of an autocorrelation $R(t_1,\, t_2) = \langle z(t_1)\, z^*(t_2)\rangle$ of $z(t)$, which is

$$\Gamma(\omega_1,\, \omega_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R(t_1,\, t_2)\, e^{-i(\omega_1 t_1 + \omega_2 t_2)}\, dt_1\, dt_2,$$

  by a formula:

$$\langle Z(\omega_1)\, Z^*(\omega_2)\rangle = \Gamma(\omega_1,\, -\omega_2). \tag{82}$$

  Proof :

  From equation (79),

$$\begin{aligned}\langle Z(\omega_1)\, Z^*(\omega_2)\rangle &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \langle z(t_1)\, z^*(t_2)\rangle\, e^{-i(\omega_1 t_1 - \omega_2 t_2)}\, dt_1\, dt_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R(t_1,\, t_2)\, e^{-i(\omega_1 t_1 - \omega_2 t_2)}\, dt_1\, dt_2 = \Gamma(\omega_1,\, -\omega_2).\end{aligned}$$

- If $z(t)$ is a stationary random process, having a power spectrum $S(\omega)$, we have

$$\langle Z(\omega_1)\, Z^*(\omega_2)\rangle = 2\pi\, S(\omega_1)\, \delta(\omega_1 - \omega_2). \tag{83}$$

Proof :

Since, in view of the stationarity of $z(t)$, its autocorrelation $\langle z(t_1)\, z^*(t_2)\rangle = R(t_1,\, t_2) = R(\tau)$ is a function of time difference $\tau = t_1 - t_2$ only. Therefore,

$$\begin{aligned}
\langle Z(\omega_1)\, Z^*(\omega_2)\rangle &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(t_1,\, t_2)\, e^{-i(\omega_1 t_1 - \omega_2 t_2)}\, dt_1\, dt_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(\tau)\, e^{-i\omega_1\tau - i(\omega_1-\omega_2)t_2}\, d\tau\, dt_2 \\
&= \int_{-\infty}^{\infty} R(\tau)\, e^{-i\omega_1\tau}\, d\tau \int_{-\infty}^{\infty} e^{-i(\omega_1-\omega_2)\, t_2}\, dt_2 \\
&= 2\pi\, S(\omega_1)\, \delta(\omega_1 - \omega_2).
\end{aligned}$$

- If $x(t)$ and $y(t)$ are jointly stationary random processes, having a cross–power spectrum $S_{xy}(\omega)$, a cross–correlation of their Fourier transforms:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-i\omega t}\, dt, \quad \text{and} \quad Y(\omega) = \int_{-\infty}^{\infty} y(t)\, e^{-i\omega t}\, dt,$$

is equal to

$$\langle X(\omega_1)\, Y^*(\omega_2)\rangle = 2\pi\, S_{xy}(\omega_1)\, \delta(\omega_1 - \omega_2). \tag{84}$$

Proof :

Since the cross–correlation $\langle x(t_1)\, y^*(t_2)\rangle = R_{xy}(t_1,\, t_2) = R_{xy}(\tau)$ is a function of time difference $\tau = t_1 - t_2$ only, we have

$$\begin{aligned}
\langle X(\omega_1)\, Y^*(\omega_2)\rangle &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{xy}(t_1,\, t_2)\, e^{-i(\omega_1 t_1 - \omega_2 t_2)}\, dt_1\, dt_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{xy}(\tau)\, e^{-i\omega_1\tau - i(\omega_1-\omega_2)t_2}\, d\tau\, dt_2 \\
&= \int_{-\infty}^{\infty} R_{xy}(\tau)\, e^{-i\omega_1\tau}\, d\tau \int_{-\infty}^{\infty} e^{-i(\omega_1-\omega_2)\, t_2}\, dt_2 \\
&= 2\pi\, S_{xy}(\omega_1)\, \delta(\omega_1 - \omega_2).
\end{aligned}$$

Thus, the autocorrelation and the cross–correlation of the Fourier transforms of the stationary random processes are uniquely related to their power and cross–power spectra by equations (83) and (84). Therefore, the Fourier transforms of the stationary random processes can be regarded as useful tools for calculating the spectra. The FX–type correlators, which first Fourier transform input records and then cross–correlate them, are the realizations of this principle.

Note that the expectation $\langle Z(\omega) \rangle$ in equation (81) has a delta–function form with respect to the angular frequency ω, that means not altogether constant in ω, except for a special case when $\langle z(t) \rangle = \eta = 0$. Also, the autocorrelation $\langle Z(\omega_1)\, Z^*(\omega_2) \rangle$ and cross–correlation $\langle X(\omega_1)\, Y^*(\omega_2) \rangle$ in equations (83) and (84) are not functions of angular–frequency difference $\omega_1 - \omega_2$ only, because of the dependence on ω_1 in $S(\omega_1)$ and $S_{xy}(\omega_1)$, except for special cases of the complete white spectra, where $S(\omega) = \text{const}$ and $S_{xy}(\omega) = \text{const}$. Therefore, the Fourier transforms of the stationary random processes are not wide–sense stationary, in general, with respect to ω.

1.3 The White Fringe

1.3.1 A Simple Interferometer

A radio interferometer, in its simplest form, can be illustrated as Figure 19.

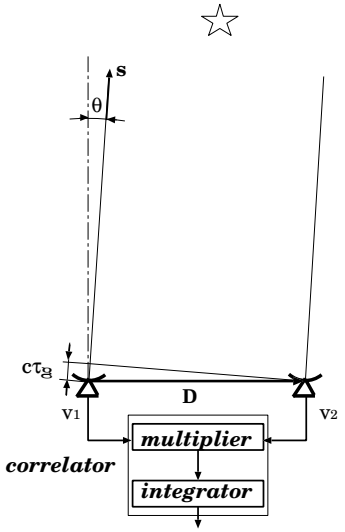

Figure 19: A simple interferometer.

This ia a two–element interferometer consisting of identical antennas, identical receivers and a correlator, which is a combination of a multiplier and an integrator (a time–averager). We ignore here the frequency conversion in the receiving system, just regarding as if the correlation processing is performed at RF (radio frequency) band.

For an infinitely distant point radio source, which we assume in this simplified case, the important quantity in the interferometry, the geometric delay τ_g, is given by

$$\tau_g = \frac{\boldsymbol{D} \cdot \boldsymbol{s}}{c} = \frac{D \sin\theta}{c},$$

as we saw in equation (1), where $\boldsymbol{D}$ is the "baseline vector" connecting reference points of two antennas, $\boldsymbol{s}$ is the "source vector" which is a unit vector directed towards the point source, c is the light velocity, and θ is an angle between the source direction $\boldsymbol{s}$ and a straight line perpendicular to the baseline vector $\boldsymbol{D}$ in the plane formed by $\boldsymbol{s}$ and $\boldsymbol{D}$. For simplicity, we assume that the same wavefront of the electromagnetic wave from an astronomical radio source are received by two antennas and then fed to a correlator with a time delay, which is equal to the geometric delay τ_g, ignoring small delays occuring in the atmosphere, transmission cables and others.

We assume a case when the beam centers of the two antennas are exactly oriented towards the radio source. Also, we assume that the source direction is close to the plane perpendicular to the baseline, i.e., $\theta \approx 0$, and hence the geometric delay τ_g is within a small range around zero. Furthermore, we ignore effects of motion of an observed radio source with respect to the baseline, due for example to the diurnal motion of the Earth, just assuming that the source is at rest or moving very slowly.

We ignore, at this stage, any contribution of the system noise, in order to concentrate our attention to the basic characteristics of the correlated radio source signals only.

In summary, we assume following properties for our simple interferometer:

- point–like radio source,
- identical antennas,
- identical receivers,
- correlation at RF–band,
- source motion is neglected,
- no delay other than the geometric,
- no system noise contribution.

1.3.2 Received Voltages as Stationary Random Processes

Let us assume that the received voltage $v(t)$, as well as the electric field intensity $E(t)$ of the radio wave which generates the voltage, are real stationary random processes, satisfying the ergodicity. Here, we used a scalar function $E(t)$ for the electric field intensity, since any antenna can receive only one polarization component of the electric field intensity vector $\boldsymbol{E}(t)$. Therefore, $E(t)$ here stands for a single polarization component of $\boldsymbol{E}(t)$ in a plane perpendicular to the direction of propagation of the transversal electromagnetic wave, which is commonly received by two antennas of the interferometer.

In actual radio astronomical observations, we usually see that the recieved voltage oscillates around zero value and its time average is just zero, i.e., time invariant. Also, the outputs of the correlators, which are time–averaged products of the received voltages, are almost constant in time, as far as we neglect the slow intrinsic time variability of the radio source. At the same time, the correlator outputs vary when we artificially insert different time delays between the two voltage time series, implying that they are functions of the time delay (i.e., of the time difference).

Therefore, the situation, which we experience in our observations, is just consistent with the stationarity and ergodicity assumptions.

Let us consider that the received voltage $v(t)$ and the electric field intensity $E(t)$, generating the voltage, are related to each other by a linear system with a real impulse response $q(t)$:

$$v(t) = E(t) * q(t) = \int_{-\infty}^{\infty} E(t-\alpha)\, q(\alpha)\, d\alpha. \tag{85}$$

Here $q(t)$ stands for the response of the antenna–receiver system to the incident radio wave, which, in particular, determines the frequency characteristics of the system as a BPF (band–pass–filter) passing a limited frequency range with a bandwidth $\Delta\omega$ centered at ω_0.

As we stated above, we assume that the responses (i.e. $q(t)$'s) of the antenna–receiver systems in the two antennas of our simple interferometer are identical, for simplicity.

1.3.3 Cross–Correlation of Received Voltages

Let us denote the received voltages of the two antennas as $v_1(t)$ and $v_2(t)$. Since they are generated by the same electromagnetic wave from a radio source, but arrived at two antennas at different times due to the geometric delay τ_g, we can express them through a common electric field intensity $E(t)$

with different time arguments as:

$$v_1(t) = \int_{-\infty}^{\infty} E(t - \tau_g - \alpha)\, q(\alpha)\, d\alpha,$$
$$v_2(t) = \int_{-\infty}^{\infty} E(t - \alpha)\, q(\alpha)\, d\alpha, \tag{86}$$

following equation (85). It is evident that $v_1(t)$ and $v_2(t)$ are jointly stationary random processes, because they are the outputs of linear systems (here, we assumed identical) with the same input stationary random process $E(t)$.

Now, let us consider that $v_1(t)$ and $v_2(t)$ are fed to the correlator shown in Figure 19. Since the correlation processing is the multiplication and integration of the signals, the correlator output $\mathcal{R}$ can be modeled as

$$\mathcal{R} = \frac{1}{\tau_a} \int_{-T}^{T} v_1(t) v_2(t) dt. \tag{87}$$

Of course, $\langle \mathcal{R} \rangle = \langle v_1(t) v_2(t) \rangle$, as far as $v_1(t)$ and $v_2(t)$ are jointly stationary random processes. Also, in view of the ergodicity, $\mathcal{R}$ tends to $\langle v_1(t) v_2(t) \rangle$ as the integration time T increases to the infinity:

$$\mathcal{R} \to \langle \mathcal{R} \rangle = \langle v_1(t) v_2(t) \rangle, \quad \text{as} \quad T \to \infty. \tag{88}$$

Therefore, assuming that the integration time is sufficiently long, we can approximate the output as

$$\mathcal{R} \cong \langle v_1(t) v_2(t) \rangle = R_{v_1 v_2}(0), \tag{89}$$

where $R_{v_1 v_2}(0)$ is the cross–correlation of the two jointly stationary random processes $v_1(t)$ and $v_2(t)$:

$$R_{v_1 v_2}(\tau) = \langle v_1(t) v_2(t - \tau) \rangle, \tag{90}$$

at zero time difference: $\tau = 0$.

Since $v_1(t)$ and $v_2(t)$ satisfy equation (86), we have

$$\begin{aligned} R_{v_1 v_2}(\tau) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \langle E(t - \tau_g - \alpha) E(t - \tau - \beta) \rangle\, q(\alpha)\, q(\beta)\, d\alpha\, d\beta \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{EE}(\tau - \tau_g - \alpha + \beta)\, q(\alpha)\, q(\beta)\, d\alpha\, d\beta, \end{aligned} \tag{91}$$

where $R_{EE}(\tau)$ is the autocorrelation of the common field $E(t)$:

$$R_{EE}(\tau) = \langle E(t)\, E(t-\tau)\rangle. \tag{92}$$

Let us introduce the cross–power spectrum $S_{v_1v_2}(\omega)$ of the received voltages $v_1(t)$ and $v_2(t)$, which forms a Fourier transform pair with the cross–correlation $R_{v_1v_2}(\tau)$, i.e., $R_{v_1v_2}(\tau) \Leftrightarrow S_{v_1v_2}(\omega)$, where ω is the angular frequency. Using Fourier transformation equation (56), the shift theorem given in equation (69), and equation (91), we obtain

$$\begin{aligned}
S_{v_1v_2}(\omega) &= \int_{-\infty}^{\infty} R_{v_1v_2}(\tau)\, e^{-i\omega\tau}\, d\tau \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{EE}(\tau - \tau_g - \alpha + \beta)\, q(\alpha)\, q(\beta)\, d\alpha\, d\beta \right] e^{-i\omega\tau}\, d\tau \\
&= \left\{ \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{EE}(\tau' - \alpha + \beta)\, q(\alpha)\, q(\beta)\, d\alpha\, d\beta \right] e^{-i\omega\tau'}\, d\tau' \right\} e^{-i\omega\tau_g} \\
&= \left[\int_{-\infty}^{\infty} R_{EE}(\tau') * q(\tau') * q(-\tau')\, e^{-i\omega\tau'}\, d\tau' \right] e^{-i\omega\tau_g}.
\end{aligned}$$

Furthermore, let us introduce a power spectrum $S_{EE}(\omega)$ of the incident electric field intensity $E(t)$, $S_{EE}(\omega) \Leftrightarrow R_{EE}(\tau)$:

$$S_{EE}(\omega) = \int_{-\infty}^{\infty} R_{EE}(\tau)\, e^{-i\omega\tau} d\tau, \tag{93}$$

and a system function $\mathcal{Q}(\omega)$ of the impulse responce $q(t)$, $\mathcal{Q}(\omega) \Leftrightarrow q(t)$:

$$\mathcal{Q}(\omega) = \int_{-\infty}^{\infty} q(t) e^{-i\omega t} dt. \tag{94}$$

Note that $S_{EE}(\omega)$ is a real and even function of ω, since $E(t)$ is a real process, and also $\mathcal{Q}^*(\omega) = \mathcal{Q}(-\omega)$ for the real impulse response $q(t)$.

Then, in view of the convolution theorem in the Fourier transformation given in equation (67), we obtain

$$S_{v_1v_2}(\omega) = S_{EE}(\omega)\, \mathcal{Q}(\omega)\, \mathcal{Q}(-\omega)\, e^{-i\omega\tau_g} = S_{EE}(\omega) \mid \mathcal{Q}(\omega) \mid^2 e^{-i\omega\tau_g}. \tag{95}$$

Inverse Fourier transformation of this equation yields the cross–correlation $R_{v_1v_2}(\tau)$:

$$R_{v_1v_2}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{v_1v_2}(\omega)\, e^{i\omega\tau} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{EE}(\omega) \mid \mathcal{Q}(\omega) \mid^2 e^{i\omega(\tau-\tau_g)} d\omega. \tag{96}$$

Taking $\tau = 0$ in this equation, we obtain

$$\mathcal{R} \cong R_{v_1 v_2}(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{EE}(\omega) \mid \mathcal{Q}(\omega) \mid^2 e^{-i\omega\tau_g} d\omega. \tag{97}$$

This is an equation which gives a relation between the expectation of the correlator output $\langle \mathcal{R} \rangle = R_{v_1 v_2}(0)$ of our simple interferometer and the spectrum of the radio wave coming from an astronomical source $S_{EE}(\omega)$, filtered by the frequency response $\mid \mathcal{Q}(\omega) \mid^2$ of the antenna–receiver systems.

As the simplest case of the frequency response $\mid \mathcal{Q}(\omega) \mid^2$, let us assume a rectangular filter:

$$\mid \mathcal{Q}(\omega) \mid^2 = \begin{cases} G & \text{if} \quad \omega_0 - \frac{\Delta\omega}{2} \leq \mid \omega \mid \leq \omega_0 + \frac{\Delta\omega}{2}, \\ 0 & \text{otherwise,} \end{cases} \tag{98}$$

where ω_0 is the band–center angular frequency, $\Delta\omega$ is the bandwidth in angular frequency, and G is the "gain" of the antenna–receiver system which is assumed here to be a constant coefficient, as shown in Figure 20.

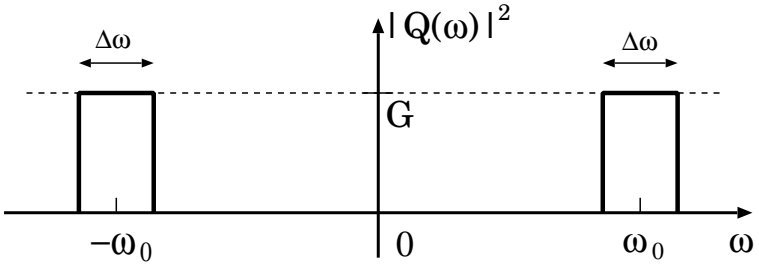

Figure 20: Rectangular frequency response of the antenna–receiver system.

On the other hand, it is usually safe to assume, for a continuum spectrum source, that the power spectrum of the radio wave $S_{EE}(\omega)$ is almost flat, or "white–noise", in the filter passband:

$$S_{EE}(\omega) = S(\omega_0) = S(-\omega_0) = \text{const.} \tag{99}$$

1.3.4 Amplitude and Phase Spectra of the Correlated Signals

In our simple interferometer, the cross–power spectrum $S_{v_1 v_2}(\omega)$ of the received voltages $v_1(t)$ and $v_2(t)$ is given by equation (95):

$$S_{v_1 v_2}(\omega) = S_{EE}(\omega) \mid \mathcal{Q}(\omega) \mid^2 \; e^{-i\omega\tau_g}.$$

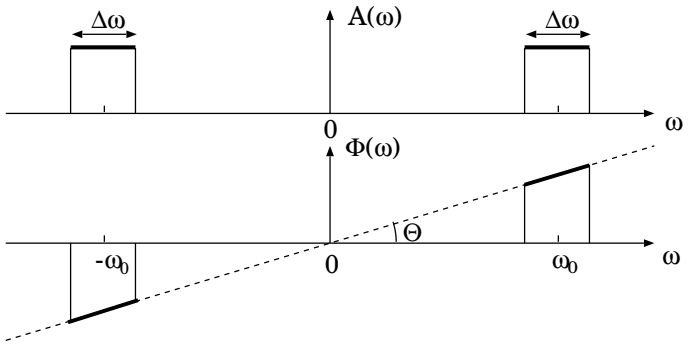

Figure 21: Amplitude (top) and phase (bottom) spectra of the cross–power spectrum.

In general, when we express a cross–power spectrum $S_{x_1x_2}(\omega)$ as

$$S_{x_1x_2}(\omega) = a(\omega)e^{-i\phi(\omega)}, \tag{100}$$

in terms of amplitude $a(\omega)$ and phase $\phi(\omega)$, we call these functions of frequency $a(\omega)$ and $\phi(\omega)$ the "amplitude spectrum" and the "phase spectrum", respectively, of the cross–power spectrum.

In the simple interferometer case, and when we assume the flat power spectrum $S_{EE}(\omega)$ (equqtion (99)) and the rectangular filter for $\mid \mathcal{Q}(\omega) \mid^2$ (equation (98)), the amplitude spectrum $A(\omega)$ and phase spectrum $\Phi(\omega)$ of the cross–power spectrum $S_{v_1v_2}(\omega)$:

$$S_{v_1v_2}(\omega) = A(\omega)e^{-i\Phi(\omega)}, \tag{101}$$

are given by

$$A(\omega) = S_{EE}(\omega) \mid \mathcal{Q}(\omega) \mid^2 \cong \text{ const}, \tag{102}$$

and

$$\Phi(\omega) = \omega\tau_g, \tag{103}$$

within the passband (see Figure 21).

Note, that the phase spectrum is expressed by a straight line, crossing the origin and tilted by an angle Θ, with an inclination equal to the geometric delay: $\tan\Theta = \tau_g$. This is a general feature of phase spectra of continuum spectrum sources which is widely seen in radio interferometer observations.

1.3.5 Correlator Output of a Band–Limited White Noise Signal

In the case of the rectangular filter and the flat spectrum source, equations (97), (98) and (99) yield a simple form for the expectation of the correlator

output $\mathcal{R}$:

$$\begin{aligned}
\langle \mathcal{R} \rangle &= \frac{1}{\pi} \Re \left[\int_0^\infty S_{EE}(\omega) \mid \mathcal{Q}(\omega) \mid^2 e^{-i\omega\tau_g} d\omega \right] \\
&= \frac{S(\omega_0)\, G}{\pi} \Re \left[\int_{\omega_0 - \frac{\Delta\omega}{2}}^{\omega_0 + \frac{\Delta\omega}{2}} e^{-i\omega\tau_g}\, d\omega \right] \\
&= \frac{S(\omega_0)\, G}{\pi} \Re \left[e^{-i\omega_0\tau_g} \int_{-\frac{\Delta\omega}{2}}^{\frac{\Delta\omega}{2}} e^{-i\omega'\tau_g}\, d\omega' \right] \\
&= 2\, B\, S(\omega_0)\, G\, \frac{\sin(\pi B \tau_g)}{\pi B \tau_g}\, \cos(\omega_0 \tau_g), \qquad (104)
\end{aligned}$$

where $\Re$ stands for the real part of a complex quantity, $B = \Delta\omega/(2\pi)$ is the frequency bandwidth, and ω' is chosen to satisfy $\omega = \omega_0 + \omega'$. In deriving

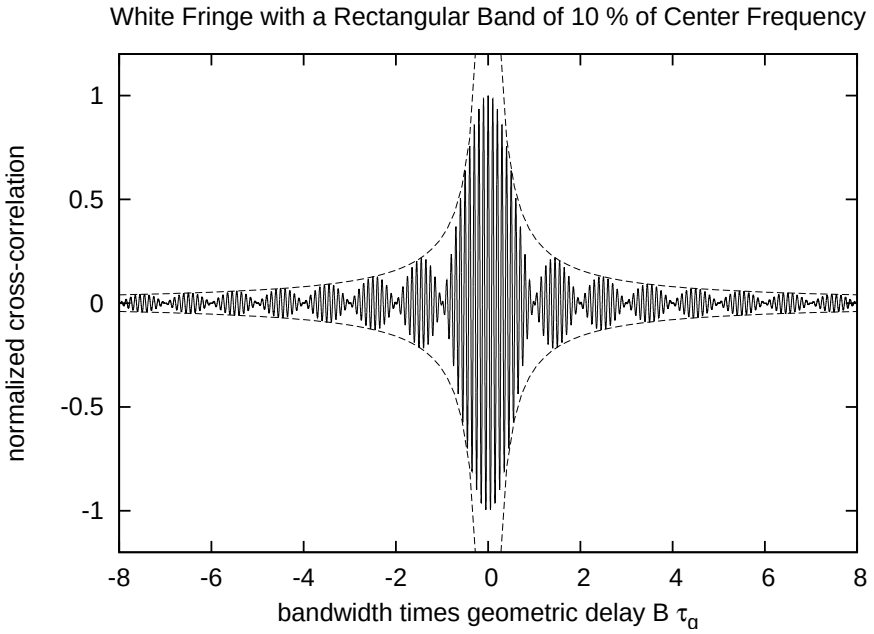

Figure 22: A normalized white fringe of a noise signal limited within a rectangular frequency band of width B, equal to 10 % of central frequency (solid line). Horizontal axis shows $B\tau_g$, i.e. the geometric delay multiplied by the bandwidth. Also shown by dashed lines is the behaviour of the $1/(\pi B\tau_g)$ term which quickly supresses the fringe amplitude with increasing τ_g.

the above equation, we used a well–known integration formula:

$$\int_{-x}^{x} e^{-ix'} dx' = \frac{e^{ix} - e^{-ix}}{i} = 2 \sin x.$$

A function having a form $\dfrac{\sin x}{x}$ is known as "sinc function".

1.3.6 Fringe Pattern Enclosed by Bandwidth Pattern

Figure (22) shows the expectation of the correlator output $\langle \mathcal{R} \rangle$ of a white noise signal from the radio source, which is limited within a rectangular passband of width B, according to equation (104). The vertical axis shows amplitude normalized by $2\,B\,S(\omega_0)\,G$, and the horizontal axis shows the geometric delay τ_g normalized by $1/B$, i.e., $B\tau_g$. In this figure, the bandwidth B is chosen to be equal to 10 % of the central frequency ($B = 0.1\omega_0/2\pi$).

We again have a fringe pattern $\cos(\omega_0 \tau_g)$ enclosed by an envelope, which, in this case, has a sinc function form, and takes the maximum value at $\tau_g = 0$. The enclosed fringe pattern, obtained from the band–limited white noise spectrum, is called the "**white fringe**".

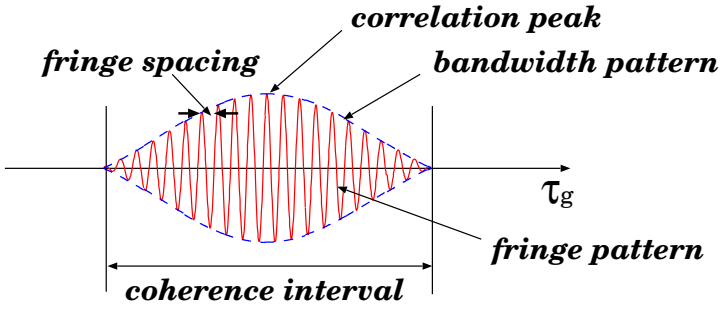

Figure 23: Technical terms describing the white fringe.

Figure 23 shows basic characteristics of the white fringe. The envelope, enclosing the rapidly oscillating fringe pattern, is called "bandwidth pattern". A particular case of the rectangular band gives the sinc function pattern, as we saw already. Other band shapes give different shapes of the bandwidth pattern. But in any case, we always have the common feature, that the interferometric fringes of finite amplitude are obtained within a limited range of the geometric delay, enclosed by a bandwidth pattern, as far as the noise signal is band–limited.

Such a limited range of the geometric delay $\Delta\tau_B$, where the fringe pattern has finite amplitude, is called "coherence interval", and is roughly represented by an equation $\Delta\tau_B = 2/B$, where B, in a general band–shape case, is a quantity which effectively characterizes a bandwidth.

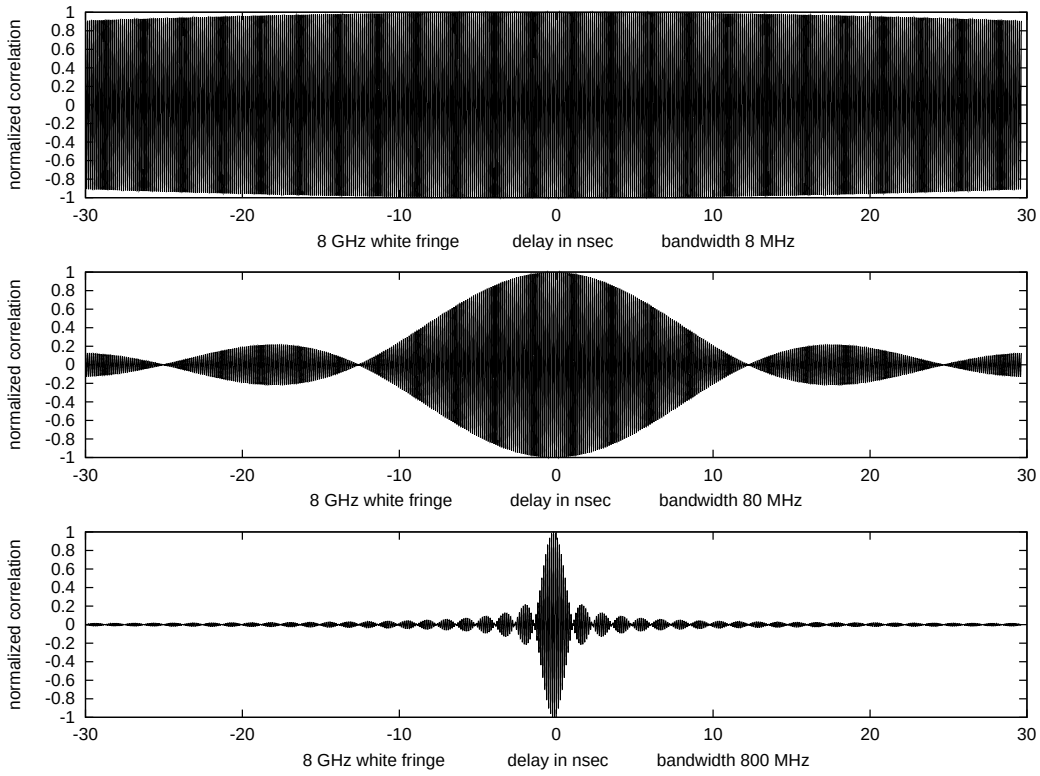

Figure 24: White fringes with ceter frequency 8 GHz and various bandwidths: $B = 8$ MHz (top), $B = 80$ MHz (middle), and $B = 800$ MHz (bottom). Horizontal axes show geometric delays covering a range from -30 to 30 nsec.

The fringe spacing (interval between successive peaks of the fringe pattern) $\Delta\tau_F$, in terms of the geometric delay, is given by a condition $\omega_0\Delta\tau_F = 2\pi$, therefore, $\Delta\tau_F = 1/\nu_0$.

The peak of the bandwidth pattern, which allows to determine a precise delay observable for the geodetic VLBI, is called "correlation peak". Of course, the peak is the sharper, the wider the bandwidth B, and therefore the narrower the coherence interval is (see Figure 24).

1.3.7 "Field of View" and "Angular Resolution" of Interferometer

A range of the geometric delay $\tau_g = D \sin\theta / c$ corresponds to an angular range in the sky. Figure 25 schematically shows the coherence interval and

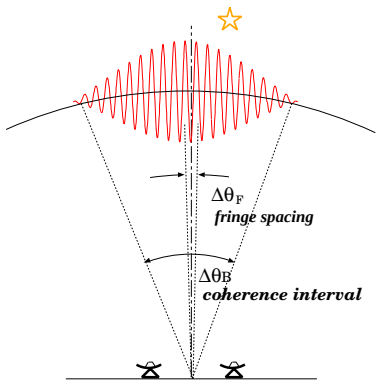

Figure 25: Coherence interval and fringe spacing in the sky.

fringe spacing in the sky.

"Field of view" of an interferometer is determined by the coherence interval. In fact, if our target source is out of the sky range corresponding to the coherence interval, we almost cannot get any signal and therefore any scientific information in our interferometry observations.

On the other hand, "angular resolution" of an interferometer is determined by the fringe spacing, as we will see later.

1.3.8 How Narrow Is the Coherence Interval in the Sky?

An angular extent $\Delta\theta_B$ corresponding to the coherence interval $\Delta\tau_B = 2/B$ in geometric delay is schematically illustrated in Figure 26.

Following the assumption which we made earlier, we consider only a region of the sky which is sufficiently close to the direction perpendicular to the baseline of an interferometer with length D, and therefore $\tau_g \simeq 0$.

For a source in the sky, which is separated from a plane perpendicular to the baseline by an angle θ, the geometric delay τ_g is equal to

$$\tau_g = \frac{D \sin\theta}{c}, \quad \text{and, hence,} \quad \theta = \arcsin\left(\frac{c\,\tau_g}{D}\right), \tag{105}$$

where c is the light velocity. Therefore, the angular extent $\Delta\theta_B$ is

$$\Delta\theta_B = 2 \arcsin\left(\frac{c\Delta\tau_B}{2D}\right) \simeq \frac{c\Delta\tau_B}{D} = \frac{2c}{DB}, \tag{106}$$

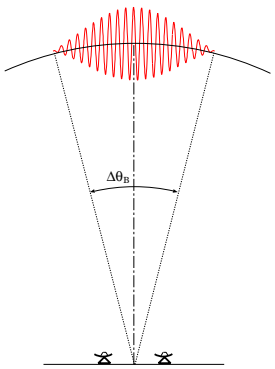

Figure 26: Coherence interval in the sky.

in the lenear approximation with respect to $c/(DB)$.

The coherence intervals for several values of D and B are listed in Table 1.

bandwidth B		200 kHz	2 MHz	200 MHz	2 GHz
$\Delta\tau_B$		10 μsec	1 μsec	10 nsec	1 nsec
$\Delta\theta_B$	$D = 100$ m	$> 180°$	$> 180°$	$1.°7$	$0.°17$
	$D = 10$ km	$17°$	$1.°7$	$1.'0$	$6.''2$
	$D = 1000$ km	$0.°17$	$1.'0$	$0.''62$	$0.''062$

Table 1: Coherence interval values for various baseline length D and bandwidth B.

It is evident from this table that the coherence interval in the sky $\Delta\theta_B$ is fairly narrow for modern interferometers, especially for VLBI. Therefore, passive observation modes, which would just "wait for" the passage of a source through the coherence interval, are extremely ineffective and unrealistic, except in "classical" systems with baselines of ~ 100 m or shorter, and bandwidths of $\sim$ a few MHz or narrower.

Consequently, modern radio interferometers are usally equipped with special means, which compensates the delay, by time–shifting one of two received signals, as we will see later.

In VLBI, it is very important to know accurate positions of the radio sources and accurate coordinates of the baseline vectors, for successful prediction and compensation of the delay, which allow us to detect the fringe within the quite narrow "field of view", the coherence interval.

1.3.9 How Narrow Is the Fringe Spacing in the Sky?

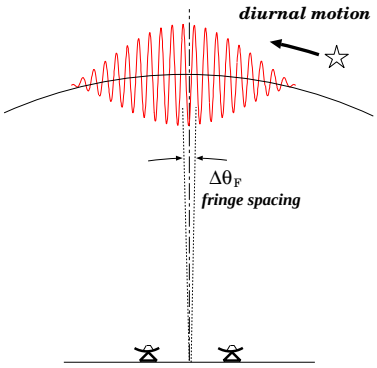

Figure 27: Fringe spacing in the sky.

Since the fringe spacing is $\Delta\tau_F = 1/\nu_0$, in terms of the geometric delay, equation (105) gives the fringe spacing in the sky $\Delta\theta_F$, in the direction nearly perpendicular to the baseline, as:

$$\Delta\theta_F = \frac{\lambda_0}{D}, \tag{107}$$

where $\lambda_0 = c/\nu_0$ is the wave length at the central frequency ν_0 of the receiving band (see Figure 27 and Table 2).

central frequency ν_0	100 MHz	10 GHz	100 GHz
wave length λ_0	3 m	3 cm	3 mm
D = 100 m	$1.^\circ7$	$1.'0$	$6.''2$
D = 10 km	$1.'0$	$0.''62$	$0.''062$
D = 1000 km	$0.''62$	$0.''0062$	$0.''00062$

Table 2: Fringe spacing values for various baseline length D and central frequency ν_0.

The angular resolution of an interferometer is usually given by equation (107), since the resolution is essentially determined by the fringe spacing, as we will discuss later.

We have so far ignored any motion of the observed source with respect to the baseline. However, every actual radio source moves in the sky due for example to the diurnal motion of the Earth. Since the source then moves

across the very dense peaks and valleys of the fringe pattern, the correlator output of our simple interferometer in equation (104) oscillates rapidly (the fringe oscillation).

In modern interferometers, this oscillation is so rapid that it causes a serious problem in actual detection of the white fringe for the following reason.

So far, we have not taken into account the system noise contribution in the received voltages, in order to focus our attention on specific features of the correlated signals from a radio source, such as the white fringe. However, in actuality, the noise inevitably exists everywhere. Moreover, in most cases, the system noise emerging in the antenna–receiver system or in the atmpsphere is much stronger than the signal from the astronomical source, as we discussed in Chapter 2 and in subsection 1.2.4.

Therefore, we usually have to integrate (time average) the correlator output for seconds, minutes, or even hours, in order to suppress the system noise effect and firmly detect the white fringe of the signal with a high enough signal–to–noise ratio. Such averaging time is far longer than the one needed to detect the white fringe in the "signal only" case when the system noise is negligibly small, as we saw in the ergodicity discussion of subsection 1.2.4. We describe this "additional integration" as "integration of the correlator output", though the actual long–time integration could be performed in the correlator itself, i.e. via long time averaging of the multiplier output by the integrator.

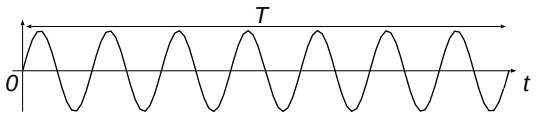

Figure 28: What will come out, if we integrate (time average) this?

Nevertheless, if the signal component in the corelator output itself is rapidly oscillating, simple averaging for a time span longer than the oscillation period would **reduce the signal to zero** (Figure 28). As a result, we never detect the white fringe of the signal above the noise level.

Therefore, modern radio interferometers are usually equipped with special means to compensate the rapid phase change, and stop the fringe oscillation, as we will see later.

2 A Realistic Radio Interferometer

2.1 Scope of the Problem

2.1.1 Limitations of the Simple Interferometer Model

The simple interferometer model, which we discussed in the previous section, was helpful for understanding one of the most important concepts in radio interferometry, the white fringe. Nevertheless, the simple model is far from real modern radio interferometers in following aspects.

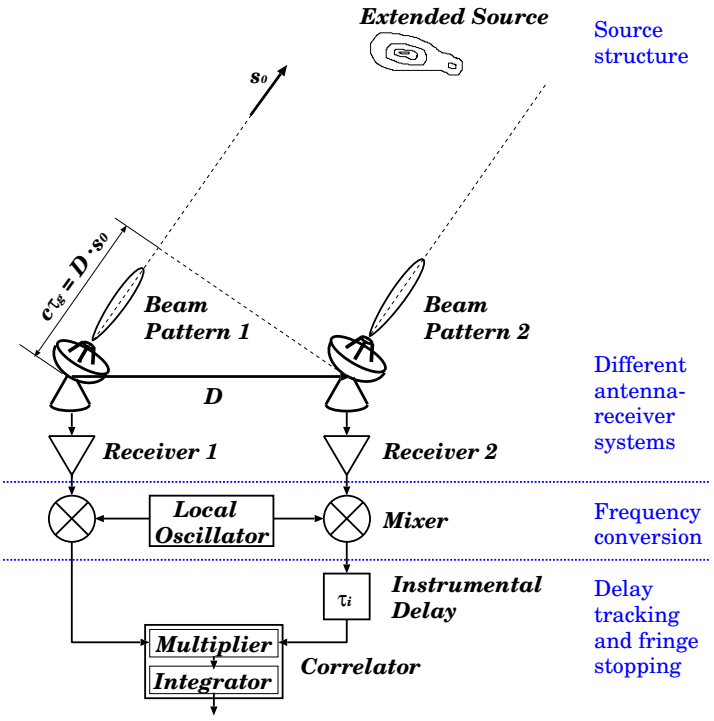

Figure 29: A realistic interferometer.

First, real radio sources are not mere points. They usually show structures, or intensity distributions, in the sky. Moreover, obtaining images of the radio source structures is one of the main purposes of VLBI, or radio interferometry in general. If the source is not point–like, the simple delayed voltage model in equation (86), which was described through a single geometric delay of a point source, is no longer valid.

Second, antennas and receiving systems in a radio interferometer are not identical in general, especially in VLBI. Therefore, beam patterns and fre-

quency responses of antenna–receiver systems may differ from each other, unlike in the simple interferometer model.

Third, receiving systems in modern interferometers are usually based on the superheterodyne design, and correlation processings are performed for IF (intermediate frequency) signals after the frequency conversion, but not for RF signals as assumed in the simple interferometer model.

Finally, the simple "wait–and–catch" type interferometer has two principal difficulties as we discussed in the last two paragraphs, namely (1) the too narrow coherence interval results in too short time for observation, and (2) the too narrow fringe spacing causes rapid fringe oscillation of the correlator output due to the motion of the source with respect to the baseline, making it almost impossible to integrate (time average) the correlator output to get high enough signal–to–noise ratio. Therefore, we have to introduce special means to compensate the geometric delay and the rapid phase change in the correlator output. Interferometry specialists call the means "**delay tracking**" and "**fringe stopping**". We will be able to observe a radio source whenever it is above the horizon, only when we successfully realize such means.

We will take into account all the above aspects in following discussions.

However, we will still neglect, for a while, small delays due for example to the atmosphere and transmission cables, taking into account only the geometric delay. Also, we will ignore, as before, any contribution of the system noise, in order to concentrate our attention to characteristics of correlated radio–source signals.

Figure 29 illustrates a "more realistic" 2–element interferometer.

2.1.2 From Radio Source to Correlator. The Way of Thinking

In a realistic interferometer, signals from a radio source pass through a series of processing units, antennas (ANT), law–noise amplifiers (LNA) and band-pass filters (BPF), down–converters (FR CONV), delay tracking (DT) and fringe stopping (FS), and then get cross–correlated (COR), as illustrated in Figure 30.

The correlator output $\mathcal{R}$, i.e. time–average of product of IF voltages $v_{F_1}(t)$ and $v_{F_2}(t)$, tends to the cross–correlation $R_{v_{F_1}v_{F_2}}(\tau)$ at $\tau = 0$:

$$\langle \mathcal{R} \rangle = R_{v_{F_1}v_{F_2}}(0) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} S_{v_{F_1}v_{F_2}}(\omega)\, d\omega,$$

with increasing integration time T (ergodicity), which is readily calculated by integrating the cross–power spectrum $S_{v_{F_1}v_{F_2}}(\omega)$ of the voltages.

Since each unit before the correlator is a kind of linear system (or filter), cross–power spectra of outputs of $i-1$–th and i–th units are related to each other through system functions of filter responses of the i–th unit.

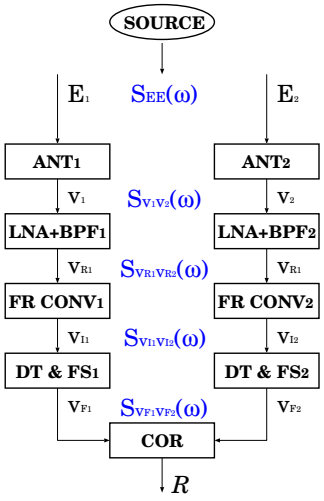

Figure 30: A realistic interferometer as a series of linear systems and a correlator.

Therefore, we can figure out the relationship of the correlator output to the incident electric field, and then to the position and structure of the source, by successively examining filter responses of all the units.

We will do this from the top to the bottom, one by one. As the first step, we will examine a relationship between the power spectrum of the incident electric field and the source structure (the intensity distribution).

2.2 Source Intensity and Incoming Electromagnetic Field

We will start our discussion on the realistic interferometer by considering how an incident electromagnetic field, coming from a radio source and being received by antennas, is related to the intensity (or brightness) distribution of the source (Figure 31).

2.2.1 Source Coherence Function

We first address ourselves to a problem regarding properties of radio waves from an extended radio source, specifically, whether radio waves coming from different points of the source are mutually correlated, or not.

Let us choose a certain direction in the source (for example, the direction of the maximum intensity), which is shown by a unit vector $\boldsymbol{s}_0$ in Figure 32,

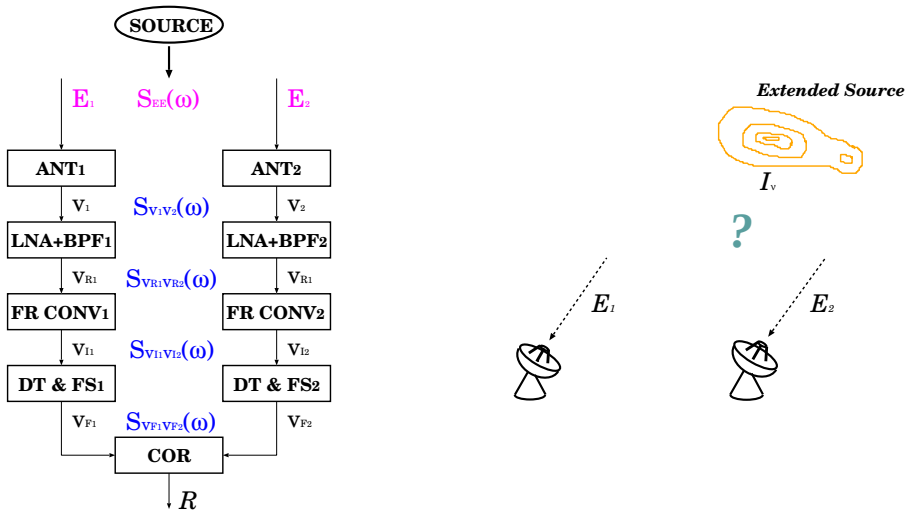

Figure 31: Relation between the intensity distribution I_ν of a radio source and the electromagnetic field E incoming from the source.

as a reference direction. Then we express a unit vector $\boldsymbol{s}$, pointing towards an arbitrary direction in the source, as $\boldsymbol{s} = \boldsymbol{s}_0 + \boldsymbol{\sigma}$. We will use the "offset vector from the reference direction" $\boldsymbol{\sigma} \equiv \boldsymbol{s} - \boldsymbol{s}_0$, as a vector indicating an element of the source at the direction $\boldsymbol{s}$. For most of radio sources with small angular extent, $\boldsymbol{\sigma}$ is almost confined in the celestial sphere.

As a quantity characterizing the electromagnetic wave from the source, we again choose a single polarization component of the electric field intensity $E(t)$, which is received by an antenna at time t. Since we now consider an extended source, we introduce a component of the electric field intensity per unit solid angle $e(\boldsymbol{\sigma}, t)$, which comes from a direction $\boldsymbol{\sigma}$ (see Figure 32). Then

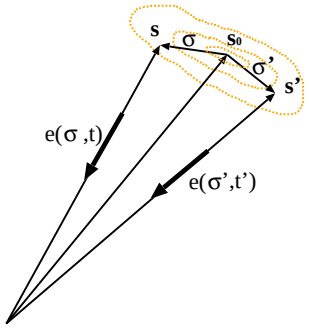

Figure 32: Radio waves coming from different directions in a radio source.

the incident electric field from the whole source $E(t)$ is given by

$$E(t) = \int\limits_{source} e(\boldsymbol{\sigma}, t) d\Omega, \tag{108}$$

where $d\Omega$ is an infinitesimal solid angle element.

Now, let us consider a cross–correlation of electric field components $e(\boldsymbol{\sigma}, t)$ and $e(\boldsymbol{\sigma}', t')$, coming from different directions $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ (or $\boldsymbol{s} = \boldsymbol{s}_0 + \boldsymbol{\sigma}$ and $\boldsymbol{s}' = \boldsymbol{s}_0 + \boldsymbol{\sigma}'$), and taken at different times t and t' (see Figure 32):

$$\langle e(\boldsymbol{\sigma}, t)\, e(\boldsymbol{\sigma}', t') \rangle.$$

Assuming again that $e(\boldsymbol{\sigma}, t)$ and $e(\boldsymbol{\sigma}', t')$ are jointly stationary random processes, we express the cross–correlation as a function of time difference $\tau = t - t'$:

$$\gamma(\boldsymbol{\sigma}, \boldsymbol{\sigma}', \tau) = \langle e(\boldsymbol{\sigma}, t)\, e(\boldsymbol{\sigma}', t') \rangle. \tag{109}$$

We call this function $\gamma(\boldsymbol{\sigma}, \boldsymbol{\sigma}', \tau)$ the "source coherence function".

Using this, we introduce a cross–correlation coefficient, as defined in equation (25), which we call the "normalized source coherence function":

$$\gamma_N(\boldsymbol{\sigma}, \boldsymbol{\sigma}', \tau) = \frac{\gamma(\boldsymbol{\sigma}, \boldsymbol{\sigma}', \tau)}{\sqrt{\gamma(\boldsymbol{\sigma}, \boldsymbol{\sigma}, 0)\gamma(\boldsymbol{\sigma}', \boldsymbol{\sigma}', 0)}}, \tag{110}$$

(here, we assume that the expectation of the electric field is zero, i.e. $\langle e(\boldsymbol{\sigma}, t) \rangle = 0$, and therefore the cross–correlation is equal to the cross–covariance). According to the general property of the cross–correlation coefficient, the normalized source coherence function always satisfies a condition:

$$0 \leq \mid \gamma_N(\boldsymbol{\sigma}, \boldsymbol{\sigma}', \tau) \mid \leq 1.$$

Now we introduce following definitions.

1. Radio waves from different directions $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ are called "completely coherent" if $\mid \gamma_N(\boldsymbol{\sigma}, \boldsymbol{\sigma}', \tau) \mid = 1$ for any τ, and "completely incoherent" if $\mid \gamma_N(\boldsymbol{\sigma}, \boldsymbol{\sigma}', \tau) \mid = 0$ for any τ.

2. A radio source is called "coherent" if radio waves from any different directions $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ in the source are completely coherent, and "incoherent" if the waves are completely incoherent. In all other cases, the radio source is called "partially coherent".

If someone puts many transmission antennas in a town, and broadcasts a TV programm, then any "poor–reception–level" problem would simply disappear. Instead, however, we would suffer from a serious "ghost" problem (many shifted images on a screen, echoed voices,...) at least in an analog TV system, since TV signals from different antennas are mutually coherent. The "ghost" images would drastically change when we slightly shift or rotate our TV reception antenna to an extent, comparable with the wavelength of the TV signal. Therefore, it would even become difficult to know directions of transmission antennas in the easiest way, i.e. by just rotating the reception antenna and watching the screen. But if the TV broadcasting is turned off, and only incoherent (independent) noises are emitted from transmission antennas, it would become much easier to know their directions.

Fortunately or not, most of astronomical radio sources are known to be incoherent, and, therefore, can be imaged relatively simply. This is because radio waves emitted from the source regions are just mutually independent noises generated by random microscopic processes occuring there. We will assume hereafter that radio sources are incoherent.

Then, the source coherence function must be expressed as

$$\gamma(\boldsymbol{\sigma}, \boldsymbol{\sigma}', \tau) = \gamma(\boldsymbol{\sigma}, \tau)\delta(\boldsymbol{\sigma} - \boldsymbol{\sigma}'), \tag{111}$$

since radio waves from directions $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ are correlated only when $\boldsymbol{\sigma} = \boldsymbol{\sigma}'$. The function $\gamma(\boldsymbol{\sigma}, \tau)$, defined by equation (111), is called the "self–coherence function".

2.2.2 Power of Electric Field Incident from a Certain Direction

Now we can express the autocorrelation $R_{EE}(\tau)$ of the incident electric field $E(t)$ given by equation (108) through the self–coherence function:

$$\begin{aligned} R_{EE}(\tau) = \langle E(t)E(t')\rangle &= \int\!\!\int_{source} \langle e(\boldsymbol{\sigma}, t)e(\boldsymbol{\sigma}', t')\rangle d\Omega d\Omega' = \int\!\!\int_{source} \gamma(\boldsymbol{\sigma}, \boldsymbol{\sigma}', \tau) d\Omega d\Omega' \\ &= \int\!\!\int_{source} \gamma(\boldsymbol{\sigma}, \tau)\delta(\boldsymbol{\sigma} - \boldsymbol{\sigma}') d\Omega d\Omega' = \int_{source} \gamma(\boldsymbol{\sigma}, \tau) d\Omega. \end{aligned} \tag{112}$$

Then, a mean square value (the "power") of the electric field $E(t)$ is given by

$$\langle E(t)^2\rangle = R_{EE}(0) = \int_{source} \gamma(\boldsymbol{\sigma}, 0) d\Omega, \tag{113}$$

which implies that $\gamma(\boldsymbol{\sigma}, 0)$ is a "solid–angle density" of the power of the single polarization component of the electric field in the direction of $\boldsymbol{\sigma}$.

Consequently, if we consider a total electric field vector $\boldsymbol{E}(t)$, which comes from a small solid angle element $\Delta\Omega$ towards a direction $\boldsymbol{\sigma}$, and includes both of two independent polarization components, then its mean square value (or power) is related to the self–coherence function at $\tau = 0$ as:

$$\frac{1}{2}\,\langle\mid \boldsymbol{E}(\boldsymbol{\sigma})\mid^2\rangle = \gamma(\boldsymbol{\sigma},0)\,\Delta\Omega, \tag{114}$$

for a randomly polarized electric field, where the coefficient $1\,/\,2$ corresponds to the fact that $\gamma(\boldsymbol{\sigma},0)$ includes only a single polarization component of the electric field.

2.2.3 Poynting Flux

In order to reveal the relation between the incident electric field and the source intensity (or brightness) distribution, we will use the Poynting vector which we discussed in Chapter 2 as a quantity describing the energy flux of the electromagnetic field.

Let us consider Poynting vector $\boldsymbol{S}(\boldsymbol{\sigma})$ of an electromagnetic wave, which comes from a small solid angle element $\Delta\Omega$ in a direction $\boldsymbol{\sigma}$. Let $\Delta\Omega$ be so small that the wave is well approximated by a superposition of monochromatic plane waves with various frequencies propagating towards the same direction $\boldsymbol{n} = -\boldsymbol{s} = -(\boldsymbol{s}_0 + \boldsymbol{\sigma})$. Let us denote m–th monochromatic plane wave component (i.e. frequency component) of the electromagnetic field quantities by suffix m.

For a monochromatic plane wave, we derived in Chapter 2 an explicit form of Poynting vector $\boldsymbol{S}_m(\boldsymbol{\sigma})$:

$$\boldsymbol{S}_m(\boldsymbol{\sigma}) = \boldsymbol{E}_m(\boldsymbol{\sigma}) \times \boldsymbol{H}_m(\boldsymbol{\sigma}) = \frac{1}{Z}\,\boldsymbol{n}\;\mid \boldsymbol{E}_m(\boldsymbol{\sigma})\mid^2, \tag{115}$$

in terms of corresponding electric and magnetic field intensities $\boldsymbol{E}_m(\boldsymbol{\sigma})$ and $\boldsymbol{H}_m(\boldsymbol{\sigma})$, respectively, where Z is the intrinsic impedance of the medium, and $\boldsymbol{n}$ is the unit vector showing the direction of propagation.

We introduce total Poynting vector $\boldsymbol{S}(\boldsymbol{\sigma})$ of the electromagnetic wave as a vector sum of all frequency components $\boldsymbol{S}_m(\boldsymbol{\sigma})$'s:

$$\boldsymbol{S}(\boldsymbol{\sigma}) = \sum_m \boldsymbol{S}_m(\boldsymbol{\sigma}) = \frac{1}{Z}\,\boldsymbol{n}\sum_m \mid \boldsymbol{E}_m(\boldsymbol{\sigma})\mid^2 .$$

Then **mean** total Poynting vector $\langle\boldsymbol{S}(\boldsymbol{\sigma})\rangle$ is given by

$$\langle\boldsymbol{S}(\boldsymbol{\sigma})\rangle = \frac{1}{Z}\,\boldsymbol{n}\sum_m\langle\mid \boldsymbol{E}_m(\boldsymbol{\sigma})\mid^2\rangle.$$

The term $\langle \mid \boldsymbol{E}_m(\boldsymbol{\sigma}) \mid^2 \rangle$ here corresponds to a power of the electric field, contained within an infinitesimally narrow frequency band. Such a power is equal to the power spectrum at the central frequency of the band, multiplied by the narrow bandwidth, as we saw in the discussion of the filterbank spectrometer in subsection 1.2.8. On the other hand, the total power of the electric field $\langle \mid \boldsymbol{E}(\boldsymbol{\sigma}) \mid^2 \rangle$, contained in the finite frequency bandwidth, is simply equal to the integral of the power spectrum over the whole bandwidth. Therefore, the total power must be just equal to a sum of powers of all monochromatic plane wave components contained in the bandwidth:

$$\langle \mid \boldsymbol{E}(\boldsymbol{\sigma}) \mid^2 \rangle = \sum_m \langle \mid \boldsymbol{E}_m(\boldsymbol{\sigma}) \mid^2 \rangle.$$

Therefore, an equation, which has the same form as the one in equation (115) for a monochromatic plane wave, must hold also for the total mean Poynting vector and the total power of the electric field with the finite bandwidth, coming from a small solid angle $\Delta\Omega$ at direction $\boldsymbol{\sigma}$:

$$\langle \boldsymbol{S}(\boldsymbol{\sigma}) \rangle = \frac{1}{Z} \boldsymbol{n} \langle \mid \boldsymbol{E}(\boldsymbol{\sigma}) \mid^2 \rangle. \tag{116}$$

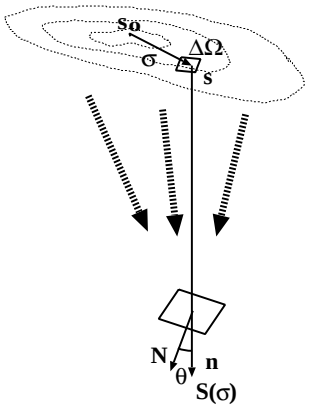

Figure 33: Poynting flux through a cross section.

Let us now introduce a new concpt which we call "Poynting flux through a cross section", or simply "Poynting flux", defined as the energy of the electromagnetic wave passing through a certain cross section of unit area, during unit duration of time (Figure 33). If we denote a unit vector normal to the cross section as $\boldsymbol{N}$, mean Poynting flux of the wave coming from the direction $\boldsymbol{\sigma}$ is given by a projection of the mean Poynting vector on $\boldsymbol{N}$:

$$\langle \boldsymbol{S}(\boldsymbol{\sigma}) \rangle \cdot \boldsymbol{N} = \frac{1}{Z} \langle \mid \boldsymbol{E}(\boldsymbol{\sigma}) \mid^2 \rangle \cos\theta, \tag{117}$$

where θ is an angle between the normal unit vector $\boldsymbol{N}$ of the cross section and the direction $\boldsymbol{n}$ of the wave propagation (see Figure 33).

Then we consider the mean Poynting flux $\langle S\rangle$ of the wave coming not only from a small solid angle towards $\boldsymbol{\sigma}$, but from the all source area:

$$\langle S\rangle = \sum_n \langle \boldsymbol{S}(\boldsymbol{\sigma}_n)\rangle \cdot \boldsymbol{N} = \frac{1}{Z}\sum_n \langle \mid \boldsymbol{E}(\boldsymbol{\sigma}_n) \mid^2\rangle \cos\theta_n, \tag{118}$$

where $\sum_n$ means a summation over all small solid angle elements $\Delta\Omega_1, \Delta\Omega_2, \cdots$, $\Delta\Omega_n, \cdots$, towards directions $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \cdots, \boldsymbol{\sigma}_n, \cdots$, correspondingly, covering the whole source area, and θ_n is an angle between the normal to the cross section $\boldsymbol{N}$ and the direction of propagation $\boldsymbol{n}_n = -(\boldsymbol{s}_0 + \boldsymbol{\sigma}_n)$.

Meanwhile, we derived in equation (114) that the power of the incident electric field is related to the self–coherence function of the radio source by

$$\frac{1}{2}\langle \mid \boldsymbol{E}(\boldsymbol{\sigma}_n) \mid^2\rangle = \gamma(\boldsymbol{\sigma}_n, 0)\Delta\Omega_n,$$

where $\Delta\Omega_n$ is a small solid angle element towards a direction $\boldsymbol{\sigma}_n$. Therefore, equations (114) and (118) yield

$$\langle S\rangle = \frac{1}{Z}\sum_n \langle \mid \boldsymbol{E}(\boldsymbol{\sigma}_n) \mid^2\rangle \cos\theta_n = \frac{2}{Z}\sum_n \gamma(\boldsymbol{\sigma}_n, 0)\cos\theta_n \Delta\Omega_n. \tag{119}$$

Replacing the summation, with respect to the small solid angles, by an integration, we obtain the relation between the mean Poyinting flux $\langle S\rangle$ and the self–coherence function $\gamma(\boldsymbol{\sigma}, \tau)$:

$$\langle S\rangle = \frac{2}{Z}\int\limits_{source} \gamma(\boldsymbol{\sigma}, 0)\cos\theta d\Omega. \tag{120}$$

Then, introducing Fourier transform $\tilde{\gamma}(\boldsymbol{\sigma}, \omega)$ of the self–coherence function $\gamma(\boldsymbol{\sigma}, \tau)$, i.e. $\gamma(\boldsymbol{\sigma}, \tau) \Leftrightarrow \tilde{\gamma}(\boldsymbol{\sigma}, \omega)$:

$$\begin{aligned} \tilde{\gamma}(\boldsymbol{\sigma}, \omega) &= \int\limits_{-\infty}^{\infty} \gamma(\boldsymbol{\sigma}, \tau)\, e^{-i\omega\tau} d\tau, \\ \gamma(\boldsymbol{\sigma}, \tau) &= \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} \tilde{\gamma}(\boldsymbol{\sigma}, \omega)\, e^{i\omega\tau} d\omega, \end{aligned} \tag{121}$$

we have

$$\gamma(\boldsymbol{\sigma}, 0) = \int\limits_{-\infty}^{\infty} \tilde{\gamma}(\boldsymbol{\sigma}, \omega)\, d\nu,$$

where $\nu = \omega/(2\pi)$ is a frequency corresponding to the angular frequency ω. Since Fourier transform of a real function is an even function of the frequency, $\gamma(\boldsymbol{\sigma}, 0)$ is reduced to

$$\gamma(\boldsymbol{\sigma}, 0) = 2 \int_0^\infty \tilde{\gamma}(\boldsymbol{\sigma}, \omega)\, d\nu. \tag{122}$$

Therefore, the mean Poynting flux in equation (120) is now given by

$$\langle S \rangle = \frac{4}{Z} \int_0^\infty \int_{source} \tilde{\gamma}(\boldsymbol{\sigma}, \omega) \cos\theta \, d\Omega \, d\nu. \tag{123}$$

2.2.4 Electric Field and Radio Source Intensity — Electromagnetics and Astronomy

As we saw in Chapter 1, results of radio astronomical observations are characterized by a number of quantities, such as, "intensity" I_ν, "spectral flux density" $\mathcal{S}_\nu$, and "power flux density" $\mathcal{S}$. These quantities are phenomenologically defined in astronomy. For example, the intensity is defined as "the quantity of radiation energy incoming from a certain direction in the sky, per unit solid angle, per unit time, per unit area perpendicular to this direction, and per unit frequency bandwidth with center frequency ν". As we see, no electromagnetic quantity, such as electric field intensity $\boldsymbol{E}$, or voltage v, appears in such a phenomenological definition. Therefore, we must precisely establish a relationship between the electromagnetic and radioastronomical quantities, in order to describe radioastronomical results in terms of the electromagnetic quantities, which we actually deal with in our radio telescopes.

Such a relationship was defined by IEEE (Institute of Electrical and Electronics Engineers) in 1977. According to the definition, the power flux density in astronomy is equal to the time average of the Poynting vector in electromagnetics.

The power flux density $\mathcal{S}$ is defined in astronomy as "the quantity of radiation energy, over the whole frequency range, incoming through a cross section of unit area, per unit time". This quantity is related to the spectral flux density $\mathcal{S}_\nu$, and to the intensity I_ν as:

$$\mathcal{S} = \int_0^\infty \mathcal{S}_\nu \, d\nu = \int_0^\infty \int_{source} I_\nu(\boldsymbol{\sigma}) \, \cos\theta \, d\Omega \, d\nu. \tag{124}$$

Since the power flux density is given with respect to a certain cross section of unit area, we interpret the "Poynting vector", in the IEEE definition as a Poynting flux, passing through the same cross section.

Then, the definition of IEEE (1977) requires that the mean Poynting flux $\langle S \rangle$ must be equal to the power flux density $\mathcal{S}$:

$$\langle S \rangle = \mathcal{S}, \tag{125}$$

since, in view of the Ergodicity, the time average must be equal to the statistical mean, provided that the averaging time is sufficiently long. In view of equations (123) and (124), (125) means

$$\int_0^\infty \int_{source} I_\nu(\boldsymbol{\sigma})\, \cos\theta \, d\Omega \, d\nu = \frac{4}{Z} \int_0^\infty \int_{source} \tilde{\gamma}(\boldsymbol{\sigma}, \omega) \cos\theta \, d\Omega \, d\nu. \tag{126}$$

Generally speaking, equal integrals do not necessarily mean equal integrands, of course. However, in our case of the stationary random signal from an incoherent source, we can equate integrands of the both sides of equation (126). In fact, the power of the electric field intensity is simply equal to a sum of contributions from all components in frequency and spatial solid angle. To make this point clearer, let us imagine a virtual source which emits radiation only in limited frequency and spatial solid–angle ranges of those of the actual radio source. The mean Poynting flux and the power flux density from this virtual source are equal to those of the actual source in the respective frequency and solid–angle ranges. They must be expressed through the self–coherence function and the intensity, just in the same forms as those given in equations (123) and (124), but with limited frequency and solid–angle ranges of the integrations. Since the definition of IEEE (1977) requires their equality, for this virtual source as well, equation (126) must hold for arbitrary frequency range $\Delta\nu$ and arbitrary spatial solid–angle range $\Delta\Omega$, i.e.,

$$\int_{\Delta\nu} \int_{\Delta\Omega} I_\nu(\boldsymbol{\sigma})\, \cos\theta \, d\Omega \, d\nu = \frac{4}{Z} \int_{\Delta\nu} \int_{\Delta\Omega} \tilde{\gamma}(\boldsymbol{\sigma}, \omega) \cos\theta \, d\Omega \, d\nu,$$

which means that the integrands must be equal to each other. Therefore,

$$I_\nu(\boldsymbol{\sigma}) = \frac{4}{Z} \tilde{\gamma}(\boldsymbol{\sigma},\, \omega). \tag{127}$$

This is a relation between the source intensity distribution and the spectrum of the self–coherence funtion, which is the solid–angle density of the power of the incident electric field. Thus, we successfully related the source intensity distribution in astronomy to the incident electric field in electromagnetics.

Note that

$$R_{EE}(\tau) = \int_{source} \gamma(\boldsymbol{\sigma}, \tau) d\Omega,$$

and hence

$$S_{EE}(\omega) = \int\limits_{source} \tilde{\gamma}(\boldsymbol{\sigma}, \omega) d\Omega.$$

Therefore, $\tilde{\gamma}(\boldsymbol{\sigma}, \omega)$ can be interpreted as a solid–angle density of the power spectrum $S_{EE}(\omega)$ of the incident electric field.

Finally, using equation (127), we obtain

$$S_{EE}(\omega) = \frac{Z}{4} \int\limits_{source} I_{\nu}(\boldsymbol{\sigma}) d\Omega. \tag{128}$$

This is the relation between the incoming electromagnetic field and the intensity (or brightness) distribution of the radio source.

2.3 Field of View of a Single–Dish Radio Telescope

Before discussing interferometry case, we will examine, as a preparatory step, how we can get source information by squaring a received voltage in a single–dish radio telescope.

2.3.1 Voltage–Field Relation

The incident electric field is converted to the voltage in a radio telescope antenna, and this voltage is processed in our detecting device (for example, a square–law detector).

In order to consider the reception process, we introduce a voltage–field relation which is described through a voltage reception pattern which characterizes the antenna beam. Since such a pattern depends on the frequency, as we saw in Chapter 2, we will use here Fourier transforms of the incident electric field $e(\boldsymbol{\sigma}, t)$ and the voltage $v(t)$. $e(\boldsymbol{\sigma}, t)$ here stands for the single polarization component of the electric field incoming from unit solid angle at a direction $\boldsymbol{\sigma}$, as before. The Fourier transforms are given by

$$\tilde{e}(\boldsymbol{\sigma}, \omega) = \int\limits_{-\infty}^{\infty} e(\boldsymbol{\sigma}, t) e^{-i\omega t} dt, \tag{129}$$

$$\tilde{v}(\omega) = \int\limits_{-\infty}^{\infty} v(t) e^{-i\omega t} dt. \tag{130}$$

Again, we assume that both $e(\boldsymbol{\sigma}, t)$ and $v(t)$ are stationary random processes.

Without being involved in details of electromagnetics, we just assume that the received voltage at a certain frequency is proportional to incident

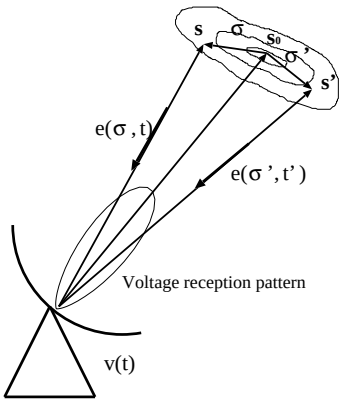

Figure 34: Voltage reception pattern.

electric field intensity at the same frequency collected by the antenna beam, or by the voltage reception pattern which we denote as $Q(\boldsymbol{\sigma}, \omega)$ (Figure 34).

Then, the voltage–field relation is given by

$$\tilde{v}(\omega) = \int\limits_{source} \tilde{e}(\boldsymbol{\sigma}, \omega) Q(\boldsymbol{\sigma}, \omega)\, d\Omega. \tag{131}$$

Note that the voltage reception pattern thus defined has a dimension of length (unit of voltage is V and unit of electric field intensity is Vm^{-1}). The voltage reception pattern is, in general, a complex quantity, since the reception process might be associated with some energy dissipation.

2.3.2 Power Pattern of a Receiving Antenna

Let us now consider a relationship between the voltage reception pattern, as defined in equation (131), and the power pattern of a receiving antenna, which we phenomenologically introduced in Chapter 2.

Introducing an autocorrelation $R_{vv}(\tau)$:

$$R_{vv}(\tau) = \langle v(t)\, v(t') \rangle, \tag{132}$$

and a correspondig power spectrum $S_{vv}(\omega)$:

$$S_{vv}(\omega) = \int\limits_{-\infty}^{\infty} R_{vv}(\tau) e^{-i\omega\tau} d\tau, \tag{133}$$

$$R_{vv}(\tau) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} S_{vv}(\omega) e^{i\omega\tau} d\omega,$$

of the received voltage $v(t)$, we describe averaged squared voltage as

$$\langle v^2(t)\rangle = R_{vv}(0) = 2\int_0^\infty S_{vv}(\omega)\, d\nu.$$

Then, the mean power W in a receiving circuit with a resistance R must be equal to the squared voltage divided by the resistance:

$$W = \int_0^\infty W_\nu\, d\nu = \frac{\langle v^2(t)\rangle}{\mathsf{R}} = \frac{R_{vv}(0)}{\mathsf{R}} = \frac{2}{\mathsf{R}}\int_0^\infty S_{vv}(\omega)\, d\nu, \tag{134}$$

where W_ν is the received power per unit bandwidth. Since this relation must hold for any segment of frequency with arbitrary bandwidth $\Delta\nu$, integrands must be equal to each other. Thus we have

$$W_\nu = \frac{2}{\mathsf{R}} S_{vv}(\omega). \tag{135}$$

Now, let us express the power spectrum $S_{vv}(\omega)$ in terms of the incident electric field intensity. For this purpose, we first calculate the autocorrelation of the Fourier transform $\tilde{v}(\omega)$ of the received voltage $v(t)$. In view of equation (131), we have

$$\langle \tilde{v}(\omega)\tilde{v}^*(\omega')\rangle = \int\!\!\int_{source} \langle \tilde{e}(\boldsymbol{\sigma},\omega)\tilde{e}^*(\boldsymbol{\sigma}',\omega')\rangle\, Q(\boldsymbol{\sigma},\omega)Q^*(\boldsymbol{\sigma}',\omega')d\Omega d\Omega'. \tag{136}$$

As we stated before, we assume that incident electric fields are stationary random processes. Therefore, the cross–correlation of their Fourier transforms $\langle \tilde{e}(\boldsymbol{\sigma},\, \omega)\tilde{e}^*(\boldsymbol{\sigma}',\, \omega')\rangle$ must satisfy the general relation given in equation (84), namely

$$\langle \tilde{e}(\boldsymbol{\sigma},\, \omega)\tilde{e}^*(\boldsymbol{\sigma}',\, \omega')\rangle = 2\pi\tilde{\gamma}(\boldsymbol{\sigma},\, \boldsymbol{\sigma}',\, \omega)\, \delta(\omega-\omega'), \tag{137}$$

where

$$\tilde{\gamma}(\boldsymbol{\sigma},\, \boldsymbol{\sigma}',\, \omega) = \int_{-\infty}^{\infty} \gamma(\boldsymbol{\sigma},\, \boldsymbol{\sigma}',\, \tau)e^{-i\omega\tau}d\tau,$$

is the cross–power spectrum of the incident electric fields from two directions $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$, or Fourier transform of the source coherence function $\gamma(\boldsymbol{\sigma},\, \boldsymbol{\sigma}',\, \tau)$, which we defined in equation (109).

For an incoherent source, the source coherence function satisfies equation (111):

$$\gamma(\boldsymbol{\sigma},\, \boldsymbol{\sigma}',\, \tau) = \gamma(\boldsymbol{\sigma},\, \tau)\, \delta(\boldsymbol{\sigma}-\boldsymbol{\sigma}'),$$

where $\gamma(\boldsymbol{\sigma}, \tau)$ is the self–coherence function. Therefore, its Fourier transform, i.e. the cross–power spectrum, must satisfy

$$\tilde{\gamma}(\boldsymbol{\sigma},\, \boldsymbol{\sigma}',\, \omega) = \tilde{\gamma}(\boldsymbol{\sigma},\, \omega)\, \delta(\boldsymbol{\sigma} - \boldsymbol{\sigma}'). \tag{138}$$

Hence, equation (137) is reduced to

$$\langle \tilde{e}(\boldsymbol{\sigma}, \omega) \tilde{e}^*(\boldsymbol{\sigma}', \omega') \rangle = 2\pi \tilde{\gamma}(\boldsymbol{\sigma}, \omega) \delta(\boldsymbol{\sigma} - \boldsymbol{\sigma}') \delta(\omega - \omega'). \tag{139}$$

Inserting this to equation (136), we obtain

$$\langle \tilde{v}(\omega) \tilde{v}^*(\omega') \rangle = 2\pi \left[\int\limits_{source} \tilde{\gamma}(\boldsymbol{\sigma}, \omega) \mid Q(\boldsymbol{\sigma}, \omega) \mid^2 d\Omega \right] \delta(\omega - \omega'). \tag{140}$$

On the other hand, according to the general property of Fourier transforms of stationary random processes as given in equation (83), the autocorrelation $\langle \tilde{v}(\omega) \tilde{v}^*(\omega') \rangle$ must be related to the power spectrum $S_{vv}(\omega)$ of the voltage as:

$$\langle \tilde{v}(\omega) \tilde{v}^*(\omega') \rangle = 2\pi S_{vv}(\omega) \delta(\omega - \omega'). \tag{141}$$

Thus, from equations (140) and (141), we obtain

$$S_{vv}(\omega) = \int\limits_{source} \tilde{\gamma}(\boldsymbol{\sigma}, \omega) \mid Q(\boldsymbol{\sigma}, \omega) \mid^2 d\Omega, \tag{142}$$

or, taking into account equations (127) and (135),

$$W_\nu = \frac{Z}{2\mathsf{R}} \int\limits_{source} I_\nu(\boldsymbol{\sigma}) \mid Q(\boldsymbol{\sigma}, \omega) \mid^2 d\Omega. \tag{143}$$

Meanwhile, in discussions of radio telescope antennas in Chapter 2, we derived an equation:

$$W_\nu = \frac{1}{2} A_e \int\limits_{source} I_\nu(\boldsymbol{\sigma}) P_n(\boldsymbol{\sigma}) d\Omega, \tag{144}$$

for the power per unit bandwidth W_ν received by an antenna, where $P_n(\boldsymbol{\sigma})$ is the normalized power pattern and A_e is the effective aperture of the antenna, and $I_\nu(\boldsymbol{\sigma})$ is the intensity distribution of the observed source.

Equation (143) must be equivalent to this equation (144). Then, noting again that we can select arbitrary solid angle as the source range, we can equate integrands of equations (143) and (144), to obtain

$$\mid Q(\boldsymbol{\sigma}, \omega) \mid^2 = \frac{\mathsf{R}}{Z} A_e P_n(\boldsymbol{\sigma}). \tag{145}$$

This is the relation between the voltage reception pattern and the normalized power pattern which determine the field of view of a radio telescope.

Now equation (142) is reduced to

$$S_{vv}(\omega) = \frac{1}{4} A_e \int\limits_{source} I_\nu(\boldsymbol{\sigma})\, P_n(\boldsymbol{\sigma})\, d\Omega, \tag{146}$$

in view of equations (127) and (145). This means that the power spectrum of the received voltage is proportional to the intensity collected by the antenna power pattern, which we called the "effective flux density" in Chapter 2 .

Thus, by squaring the received voltage, we confirmed once more that a single–dish radio telescope is a "huge power meter" indicating the effective flux density, and thus capable of revealing source image by scanning its beam.

2.3.3 New Dimensions of Voltage and Electric Field

For further discussions, it is convenient to redefine the volatage and the electric field intensity as

$$\begin{aligned} \frac{v(t)}{\sqrt{\mathsf{R}}} &\longrightarrow v(t), \\ \frac{e(\boldsymbol{\sigma},\, t)}{\sqrt{Z}} &\longrightarrow e(\boldsymbol{\sigma},\, t), \end{aligned} \tag{147}$$

to eliminate constant coefficients of the resistance R and the intrinsic impedance Z, which may cause rather complicated appearances of our equations. Then, dimensions of the voltage and the electric field intensity change to

- voltage v [electric power]$^{1/2}$ (W$^{1/2}$)
- electric field intensity e [power flux density]$^{1/2}$ (W$^{1/2}$ m^{-1}),

and we have

$$\begin{aligned} \frac{S_{vv}(\omega)}{\mathsf{R}} &\longrightarrow S_{vv}(\omega), \\ \frac{\tilde{\gamma}(\boldsymbol{\sigma},\omega)}{Z} &\longrightarrow \tilde{\gamma}(\boldsymbol{\sigma},\omega), \end{aligned} \tag{148}$$

for the power spectra of the voltage and the electric field intensity.

Let us further redefine the voltage reception pattern as

$$\sqrt{\frac{Z}{\mathsf{R}}} Q(\boldsymbol{\sigma},\omega) \longrightarrow Q(\boldsymbol{\sigma},\omega). \tag{149}$$

Then, the voltage — field relation remain in the same form as equation (131):

$$\tilde{v}(\omega) = \int\limits_{source} \tilde{e}(\boldsymbol{\sigma}, \omega) Q(\boldsymbol{\sigma}, \omega) d\Omega,$$

and we obtain simple expressions:

$$\begin{aligned} W_\nu &= 2S_{vv}(\omega), \\ I_\nu(\boldsymbol{\sigma}) &= 4\tilde{\gamma}(\boldsymbol{\sigma}, \omega), \\ A_e P_n(\boldsymbol{\sigma}) &= \mid Q(\boldsymbol{\sigma}, \omega) \mid^2, \end{aligned} \tag{150}$$

for the power per unit bandwidth, the source intensity, and the normalized power pattern. We will use these new definitions in further discussions of radio interferometry.

2.4 How Does an Interferometer View the Universe?

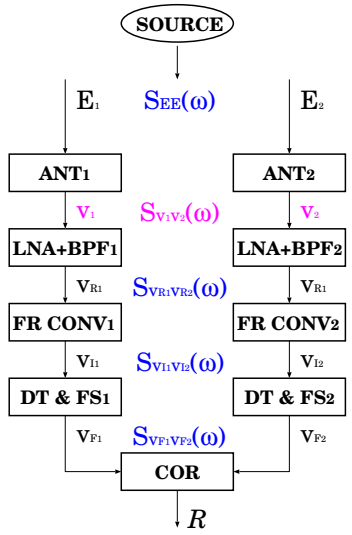

Figure 35: Cross–power spectrum of "just received" voltages.

We consider now a cross–power spectrum of voltages "just received" by antennas, and discuss how the cross–power spectrum is related to the intensity (or brightness) distribution of the source (Figure 35).

This will also be an important step towards getting an answer to a fundamental question of radio interferometry: how we can obtain an image of a radio source by cross–correlating received voltages (Figure 36).

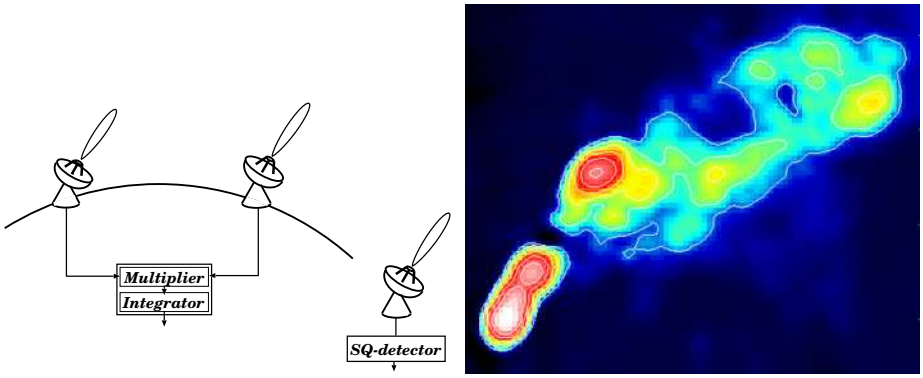

Figure 36: A fundamental question on radio interferometry. In case of a single–dish radio telescope with a square–law (SQ) detector, we can regard the telescope as a "huge power meter" detecting a power coming from a beam direction. Therefore, it is intuitively easy to understand that, by orienting the telescope beam towards different directions of the sky, we can obtain distribution of the power from a radio source, i.e. its intensity (or brightness) distribution. On the other hand, it is not intuitively clear, at least at first, why we can obtain the intensity distribution of a source, such as shown in left panel, by cross–correlating voltages received from the source at different antennas. Left panel is a 5 GHz VLBI image of a quasar 3C 380 (figure courtesy of EVN).

2.4.1 Geometric Delay of a Point in an Extended Source

Let us again consider a two–element interferometer, such as shown in Figure 37. Since we now consider an extended source, we no longer have a single geometric delay, such as the one we had for a point source. Instead, we introduce a concept of a geometric delay of a point in the extended source. Let a baseline vector $\boldsymbol{D}$ be drawn from antenna 1 to antenna 2, and let a unit vector $\boldsymbol{s} = \boldsymbol{s}_0 + \boldsymbol{\sigma}$ point towards a certain direction within a radio source, where $\boldsymbol{s}_0$ shows a reference direction of the source, as before.

A geometric delay of the point $\boldsymbol{s}$, i.e. the time interval between arrivals of the same wave front from the direction $\boldsymbol{s}$ at two antennas, is given by

$$\tau_g(\boldsymbol{\sigma}) = \frac{\boldsymbol{D} \cdot \boldsymbol{s}}{c} = \tau_{g_0} + \frac{\boldsymbol{D} \cdot \boldsymbol{\sigma}}{c}, \tag{151}$$

where c is the light velocity, $\tau_{g_0} \equiv \dfrac{\boldsymbol{D} \cdot \boldsymbol{s}_0}{c}$ is a geometric delay of the wave coming from the reference direction $\boldsymbol{s}_0$, and $\boldsymbol{\sigma} = \boldsymbol{s} - \boldsymbol{s}_0$ is the offset vector showing the direction $\boldsymbol{s}$ within the source. As we stated before, we ignore here any atmospheric or instrumental delay, other than the geometric delay.

Now, the electric field intensities $e_1(\boldsymbol{\sigma}, t)$ and $e_2(\boldsymbol{\sigma}, t)$, with a single polarization component, which arrive at two antennas at a certain time t from

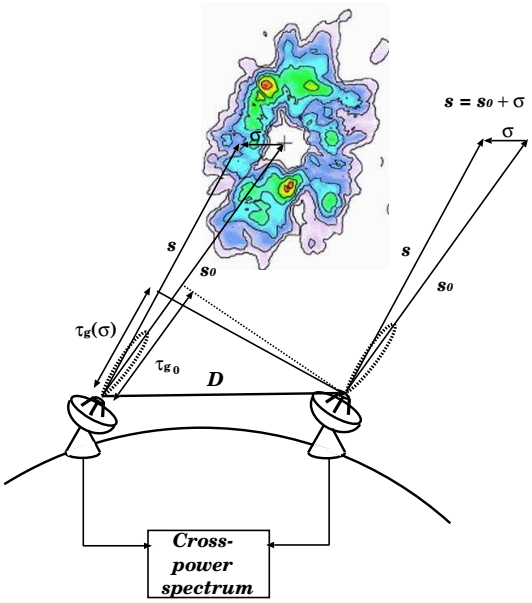

Figure 37: Geometry of a two element radio interferometer.

the same direction $\boldsymbol{\sigma}$, are described through a common incident field intensity $e(\boldsymbol{\sigma}, t)$ representing the same wave front, as

$$\begin{aligned} e_1(\boldsymbol{\sigma}, t) &= e(\boldsymbol{\sigma}, t - \tau_g(\boldsymbol{\sigma})), \\ e_2(\boldsymbol{\sigma}, t) &= e(\boldsymbol{\sigma}, t). \end{aligned} \tag{152}$$

Since we assume that $e(\boldsymbol{\sigma}, t)$ is a stationary random process, $e_1(\boldsymbol{\sigma}, t)$ and $e_2(\boldsymbol{\sigma}, t)$ are jointly stationary random processes. Then, their Fourier transforms $\tilde{e}_1(\boldsymbol{\sigma}, \omega)$ and $\tilde{e}_2(\boldsymbol{\sigma}, \omega)$ are given by

$$\begin{aligned} \tilde{e}_1(\boldsymbol{\sigma}, \omega) &= \tilde{e}(\boldsymbol{\sigma}, \omega) e^{-i\omega\tau_g(\boldsymbol{\sigma})}, \\ \tilde{e}_2(\boldsymbol{\sigma}, \omega) &= \tilde{e}(\boldsymbol{\sigma}, \omega), \end{aligned} \tag{153}$$

where $\tilde{e}(\boldsymbol{\sigma}, \omega)$ is the Fourier transform of $e(\boldsymbol{\sigma}, t)$. We used the shift theorem given in equation (69), to derive the upper one of equation (153).

2.4.2 Cross–Power Spectrum of Received Voltages

The electric field intensities, coming from the whole area of the source, are converted to voltages at the two antennas according to the voltage–field re-

lation which we showed in equation (131), i.e.

$$\tilde{v}_1(\omega) = \int\limits_{source} \tilde{e}_1(\boldsymbol{\sigma}, \omega) Q_1(\boldsymbol{\sigma}, \omega) d\Omega,$$
$$\tilde{v}_2(\omega) = \int\limits_{source} \tilde{e}_2(\boldsymbol{\sigma}, \omega) Q_2(\boldsymbol{\sigma}, \omega) d\Omega. \tag{154}$$

Therefore, their cross–correlation is given by

$$\langle \tilde{v}_1(\omega)\, \tilde{v}_2^*(\omega') \rangle = \int\limits_{source} \int\limits \langle \tilde{e}_1(\boldsymbol{\sigma},\, \omega) \tilde{e}_2^*(\boldsymbol{\sigma}',\, \omega') \rangle Q_1(\boldsymbol{\sigma},\, \omega) Q_2^*(\boldsymbol{\sigma}',\, \omega') d\Omega d\Omega'. \tag{155}$$

In view of the delayed common field nature of the incident fields, we can describe $\tilde{e}_1(\boldsymbol{\sigma},\, \omega)$ and $\tilde{e}_2(\boldsymbol{\sigma}',\, \omega')$ through the common field $\tilde{e}(\boldsymbol{\sigma},\, \omega)$ as given in equation (153). Thus, their cross–correlation is reduced to

$$\langle \tilde{e}_1(\boldsymbol{\sigma},\, \omega) \tilde{e}_2^*(\boldsymbol{\sigma}',\, \omega') \rangle = \langle \tilde{e}(\boldsymbol{\sigma},\, \omega) \tilde{e}^*(\boldsymbol{\sigma}',\, \omega') \rangle e^{-i\omega\tau_g(\boldsymbol{\sigma})}. \tag{156}$$

As we saw in equation (139),

$$\langle \tilde{e}(\boldsymbol{\sigma},\, \omega) \tilde{e}^*(\boldsymbol{\sigma}',\, \omega') \rangle = 2\pi \tilde{\gamma}(\boldsymbol{\sigma},\, \omega)\, \delta(\boldsymbol{\sigma} - \boldsymbol{\sigma}')\, \delta(\omega - \omega'),$$

where $\tilde{\gamma}(\boldsymbol{\sigma},\, \omega)$ is the Fourier transform of the self–coherence function of the incident field, since we assume that the electric field is a stationary random process and the radio source is incoherent. Therefore, equation (155) is now reduced to

$$\langle \tilde{v}_1(\omega)\, \tilde{v}_2^*(\omega') \rangle = 2\pi \left[\int\limits_{source} \tilde{\gamma}(\boldsymbol{\sigma},\, \omega)\, Q_1(\boldsymbol{\sigma},\, \omega) Q_2^*(\boldsymbol{\sigma},\, \omega) e^{-i\omega\tau_g(\boldsymbol{\sigma})}\, d\Omega \right] \delta(\omega - \omega'). \tag{157}$$

On the other hand, since voltages received at two antennas are jointly stationary random processes, the cross–correlation of their Fourier transforms and their cross–power spectrum are related to each other by the general formula given in equation (84):

$$\langle \tilde{v}_1(\omega)\, \tilde{v}_2^*(\omega') \rangle = 2\pi S_{v_1 v_2}(\omega)\, \delta(\omega - \omega'). \tag{158}$$

Then, from equations (157) and (158), we obtain

$$S_{v_1 v_2}(\omega) = \int\limits_{source} \tilde{\gamma}(\boldsymbol{\sigma},\, \omega)\, Q_1(\boldsymbol{\sigma},\, \omega)\, Q_2^*(\boldsymbol{\sigma},\, \omega)\, e^{-i\omega\tau_g(\boldsymbol{\sigma})}\, d\Omega. \tag{159}$$

This is a relationship between a cross–power spectrum of received voltages from a source and a self–coherence function of the same source. Also, this is

an interferometer analog of the single–dish relationship, which we derived in equation (142):

$$S_{vv}(\omega) = \int\limits_{source} \tilde{\gamma}(\boldsymbol{\sigma}, \omega) \mid Q(\boldsymbol{\sigma}, \omega) \mid^2 d\Omega.$$

2.4.3 Normalized Power Pattern of an Interferometer

In view of equation (150), we know that the Fourier transform $\tilde{\gamma}(\boldsymbol{\sigma}, \omega)$ of the self–coherence function and the source intensity distribution $I_\nu(\boldsymbol{\sigma})$ are related to each other by an equation:

$$\tilde{\gamma}(\boldsymbol{\sigma}, \omega) = \frac{1}{4} I_\nu(\boldsymbol{\sigma}).$$

Furthermore, equation (150) also shows that the voltage reception pattern $Q(\boldsymbol{\sigma}, \omega)$ and the normalized power pattern $P_n(\boldsymbol{\sigma})$ of a single–dish radio telescope are related to each other as

$$A_e \, P_n(\boldsymbol{\sigma}) = \mid Q(\boldsymbol{\sigma}, \omega) \mid^2,$$

where A_e is the effective aperture of the antenna. So, let us introduce here an analog of this equation in our interferometer case:

$$A_0 \, A_N(\boldsymbol{\sigma}) = Q_1(\boldsymbol{\sigma}, \omega) \, Q_2^*(\boldsymbol{\sigma}, \omega), \tag{160}$$

where A_0 is a geometrical mean of effective apertures of two antennas:

$$A_0 = \sqrt{A_{e_1} \, A_{e_2}}, \tag{161}$$

and $A_N(\boldsymbol{\sigma})$ is a quantity which we call "normalized power pattern of an interferometer".

When the two voltage reception patterns Q_1 and Q_2 are real quantities (i.e., there is no phase change associated with the field to voltage conversion), or they have equal phases, $A_N(\boldsymbol{\sigma})$ is described through the normalized power patterns P_{n1} and P_{n2} of the two antennas as:

$$A_N(\boldsymbol{\sigma}) = \sqrt{P_{n_1}(\boldsymbol{\sigma}) \, P_{n_2}(\boldsymbol{\sigma})}. \tag{162}$$

Now, taking into account equations (150), (151) and (160), and denoting

$$\omega \, \tau_g(\boldsymbol{\sigma}) = \omega \, \tau_{g_0} + \omega \, \frac{\boldsymbol{D} \cdot \boldsymbol{\sigma}}{c} = \omega \, \tau_{g_0} + 2\pi \boldsymbol{D}_\lambda \cdot \boldsymbol{\sigma}, \tag{163}$$

where $\boldsymbol{D}_\lambda = \boldsymbol{D}/\lambda$ is the baseline vector normalized by the wavelength $\lambda = 2\pi c/\omega$, we can rewrite equation (159) in a form:

$$S_{v_1 v_2}(\omega) = \frac{1}{4} A_0 \, e^{-i\omega\tau_{g_0}} \int\limits_{source} A_N(\boldsymbol{\sigma}) \, I_\nu(\boldsymbol{\sigma}) \, e^{-i2\pi \boldsymbol{D}_\lambda \cdot \boldsymbol{\sigma}} \, d\Omega. \qquad (164)$$

This equation gives a relationship between the cross–power spectrum of the received voltages and the intensity distribution of the radio source. This is an interferometer analog of the power spectrum – intensity relationship in a single dish radio telescope, which we obtained in equation (146):

$$S_{vv}(\omega) = \frac{1}{4} A_e \int\limits_{source} P_n(\boldsymbol{\sigma}) \, I_\nu(\boldsymbol{\sigma}) \, d\Omega.$$

2.4.4 Complex Visibility

We can formally expand the range of integration in equation (164) to the whole sky, since $I_\nu(\boldsymbol{\sigma}) = 0$ beyond the edge of the source, and, for sources far from the antenna beam, $A_N(\boldsymbol{\sigma}) = 0$. Then, the cross–power spectrum of the received signals is given by

$$S_{v_1 v_2}(\omega) = \frac{1}{4} A_0 \, e^{-i\omega\tau_{g_0}} \oint A_N(\boldsymbol{\sigma}) \, I_\nu(\boldsymbol{\sigma}) \, e^{-i2\pi \boldsymbol{D}_\lambda \cdot \boldsymbol{\sigma}} \, d\Omega.$$

We call the integral in this equation **"complex visibility"**, or "fringe visibility", or just "visibility", and denote it as $\mathcal{V}(\omega)$, i.e.

$$\mathcal{V}(\omega) = \oint A_N(\boldsymbol{\sigma}) \, I_\nu(\boldsymbol{\sigma}) \, e^{-i2\pi \boldsymbol{D}_\lambda \cdot \boldsymbol{\sigma}} \, d\Omega. \qquad (165)$$

We can describe the complex visibility through its amplitude $\mid \mathcal{V}(\omega) \mid$ and phase Φ_v as

$$\mathcal{V}(\omega) = \mid \mathcal{V}(\omega) \mid e^{i\Phi_v(\omega)}. \qquad (166)$$

The amplitude $\mid \mathcal{V}(\omega) \mid$ is called "visibility amplitude" or "correlated flux density", and the phase Φ_v is called "visibility phase".

The complex visibility as shown in equation (165) has a dimension of the flux density (w m^{-2} Hz^{-1}). In fact, the complex visibility is an interferometer analog of the effective flux density $\mathcal{S}_\nu$ received by a single–dish antenna

$$\mathcal{S}_\nu = \oint P_n(\boldsymbol{\sigma}) \, I_\nu(\boldsymbol{\sigma}) \, d\Omega, \qquad (167)$$

which we introduced in Chapter 2, where $P_n(\boldsymbol{\sigma})$ is the normalized power pattern of the antenna as before. This is why the visibility amplitude is also called "correlated flux density".

Although equations (165) and (167) look similar, there is a big difference as well. While equation (167) shows that the effective flux density is an "intensity collected by an antenna beam" $P_n(\boldsymbol{\sigma})$, equation (165) shows something different in view of the existence of $e^{-i2\pi \boldsymbol{D}_\lambda \cdot \boldsymbol{\sigma}}$ term. This term represents the fringe pattern of the interferometer, and usually (at least in modern interferometers) varies with $\boldsymbol{\sigma}$ much more rapidly than the normalized power pattern of an interferometer $A_N(\boldsymbol{\sigma})$ also figuring in equation (165). Therefore, we can regard the complex visibility as an "intensity collected by a fringe pattern" of an interferometer.

The cross–power spectrum of the received voltages is given through the complex visibility as:

$$S_{v_1 v_2}(\omega) = \frac{1}{4} A_0 \, e^{-i\omega\tau_{g_0}} \, \mathcal{V}(\omega). \tag{168}$$

This is an interferometer analog of the single–dish relation:

$$S_{vv}(\omega) = \frac{1}{2} W_\nu = \frac{1}{4} A_e \, \mathcal{S}_\nu, \tag{169}$$

where A_e is the effective aperture of the antenna, and W_ν is the received power per unit frequency, as before.

The complex visibility $\mathcal{V}(\omega)$ is the fundamental quantity for "imaging" (or infering structure of) an observed source, and the geometric delay τ_{g_0} of a reference direction of the source is the fundamental quantity for geodetic VLBI, as we will discuss later. Note that they are both contained in the cross–power spectrum of the received voltages as shown in equation (168). We can say that a radio interferometer views the universe through the complex visibility and the geometric delay.

2.4.5 "Quasi–Static Approach"

In the cross–power spectrum of the received voltages shown in equation (168), both the geometric delay τ_{g_0} and the complex visibility $\mathcal{V}(\omega)$ usually vary in time t, because the baseline vector $\boldsymbol{D}$ changes its direction with respect to the source direction $\boldsymbol{s}$ due for example to the diurnal rotation of the Earth.

Therefore, strictly speaking, the condition for the stationary random process is violated here, because statistical quantities, the cross–power spectrum in the present case, depends on time.

In actuality, we implicitly introduced this inconsistency when we assumed that received voltages in equation (86) for the simple interferometer, and also incident electric fields in equation (152) for the realistic interferometer,

are jointly stationary random processes, ignoring the time variability of the geometric delay.

However, we will keep using formulae for the stationary random processes in further discussions, simply assuming that the time variation is slow enough for applying the concepts of the stationary random processes.

We will use this "quasi–static approach" whenenver we will discuss time variable correlations of our signals due to clock–offset, atmospheric refraction effect, cable extension, as well as delay tracking and fringe stopping.

2.5 Response of RF Filter

The "just received" signals are usually amplified and band–pass filtered, before the frequency conversion, in cm– and long mm–wave observations. Therefore, we now consider responses of the RF filters composed of LNAs (low noise amplifiers) and BPFs (band–pass filters) (Figure 38). In cases, when first stage devices in receiving systems are mixers, such as SIS mixers used in short mm–wave observations, this step must be simply ommited.

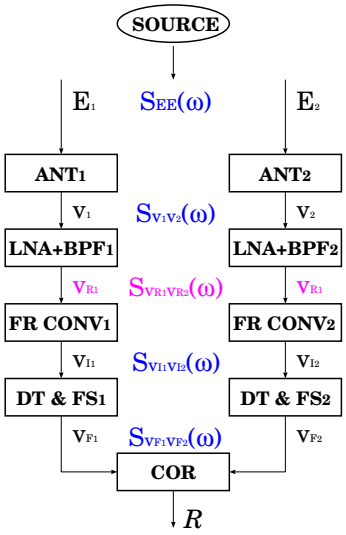

Figure 38: Cross–power spectrum of RF voltages.

2.5.1 RF Filter as a Linear System

Let us model an RF filter by a linear system, whose input is the "just received" voltage $v(t)$, output is an RF voltage $v_R(t)$, and filter response is described by a real impulse response $h_R(t)$, as illustrated in Figure 39.

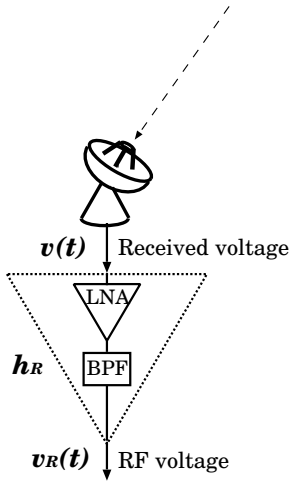

Figure 39: An image of an RF filter.

Then, we have

$$v_R(t) = v(t) * h_R(t). \tag{170}$$

In view of the convolution theorem given in equation (67), Fourier transforms $\tilde{v}(\omega)$ and $\tilde{v}_R(\omega)$ of $v(t)$ and $v_R(t)$, which we assume to be stationary random processes, and system function $H_R(\omega)$ of the impulse response $h_R(t)$:

$$v(t) \Leftrightarrow \tilde{v}(\omega), \quad v_R(t) \Leftrightarrow \tilde{v}_R(\omega), \quad h_R(t) \Leftrightarrow H_R(\omega), \tag{171}$$

satisfy

$$\tilde{v}_R(\omega) = \tilde{v}(\omega)\, H_R(\omega). \tag{172}$$

The system function $H_R(\omega)$ here stands for effects of gain in the LNA and band–pass characteristics of the BPF.

2.5.2 Cross–Power Spectrum of RF Voltages

Thus, for Fourier transforms of input voltages $v_1(t)$ and $v_2(t)$, received at two antennas, and output voltages $v_{R1}(t)$ and $v_{R2}(t)$ from RF filters, with impulse responses $h_{R1}(t)$ and $h_{R2}(t)$:

$$\begin{aligned} &v_1(t) \Leftrightarrow \tilde{v}_1(\omega), \quad v_{R1}(t) \Leftrightarrow \tilde{v}_{R1}(\omega), \quad h_{R1}(t) \Leftrightarrow H_{R1}(\omega), \\ &v_2(t) \Leftrightarrow \tilde{v}_2(\omega), \quad v_{R2}(t) \Leftrightarrow \tilde{v}_{R2}(\omega), \quad h_{R2}(t) \Leftrightarrow H_{R2}(\omega), \end{aligned} \tag{173}$$

we have

$$\begin{aligned} \tilde{v}_{R1}(\omega) &= \tilde{v}_1(\omega)\, H_{R1}(\omega), \\ \tilde{v}_{R2}(\omega) &= \tilde{v}_2(\omega)\, H_{R2}(\omega). \end{aligned} \tag{174}$$

Therefore, cross–correlation of $\tilde{v}_{R1}(\omega)$ and $\tilde{v}_{R2}(\omega)$ is givn by

$$\langle \tilde{v}_{R1}(\omega)\, \tilde{v}^*_{R2}(\omega')\rangle = \langle \tilde{v}_1(\omega)\, \tilde{v}^*_2(\omega')\rangle\, H_{R1}(\omega)\, H^*_{R2}(\omega'). \tag{175}$$

Now, in view of the general properties of cross–correlations of Fourier transforms of stationary random processes, given in equation (84), we have

$$\begin{aligned} \langle \tilde{v}_{R1}(\omega)\, \tilde{v}^*_{R2}(\omega')\rangle &= 2\pi\, S_{v_{R1}v_{R2}}(\omega)\, \delta(\omega-\omega'), \\ \langle \tilde{v}_1(\omega)\, \tilde{v}^*_2(\omega')\rangle &= 2\pi\, S_{v_1v_2}(\omega)\, \delta(\omega-\omega'), \end{aligned} \tag{176}$$

where $S_{v_1v_2}(\omega)$ and $S_{v_{R1}v_{R2}}(\omega)$ are cross–power spectra of input voltages $v_1(t)$ and $v_2(t)$, and output voltages $v_{R1}(t)$ and $v_{R2}(t)$, respectively, of the RF filters. Consequently, equation (175) yields a relation:

$$S_{v_{R1}v_{R2}}(\omega) = S_{v_1v_2}(\omega)\, H_{R1}(\omega)\, H^*_{R2}(\omega), \tag{177}$$

between the cross–power spectra of the input and output voltages.

Note that the cross–power spectrum of the received voltages $S_{v_1v_2}(\omega)$ is related to the complex visibility of an observed source by equation (168).

2.6 Frequency Conversion in Radio Interferometers

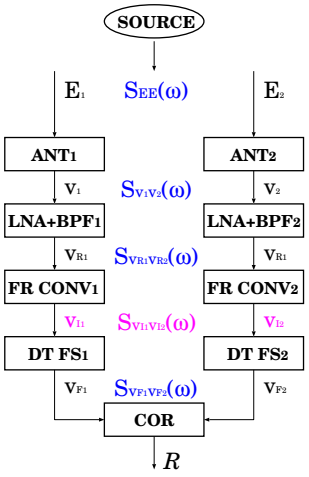

Figure 40: A relation between a cross–power spectrum $S_{v_{I1}v_{I2}}(\omega)$ of IF voltages and a cross-power spectrum $S_{v_{R1}v_{R2}}(\omega)$ of input RF voltages.

It has been technically difficult to cross–correlate voltages at RF frequencies higher than $\sim$ 1 GHz. Therefore, in modern radio interferometers, the

RF voltages are usually frequency–converted to lower IF frequencies before the correlation processing.

So, let us now consider how cross–power spectra $S_{v_{I1}v_{I2}}(\omega)$ of IF voltages after the frequency conversion are related to the cross–power spectra $S_{v_{R1}v_{R2}}(\omega)$ of RF voltages (Figure 40).

It is interesting to examine, in particular, if the condition of stationary randomness of the RF voltages is conserved after the frequency conversion.

2.6.1 Fourier Transform of IF Voltage Signal

We often perform an actual frequency conversion through several separate conversions (a multi–step conversion scheme), sequentially shifting a spectral range containing observed data from higher frequencies to lower frequencies, step by step. However, we assume here a simplified single–step conversion scheme where RF voltage is down–converted to final IF voltage at once. Basically, a single–step conversion is equivalent to a multi–step conversion, as long as total frequency shift and total band–pass characteristics are the same in both schemes. Therefore, our single–step assumption will not lower generality of following discussions.

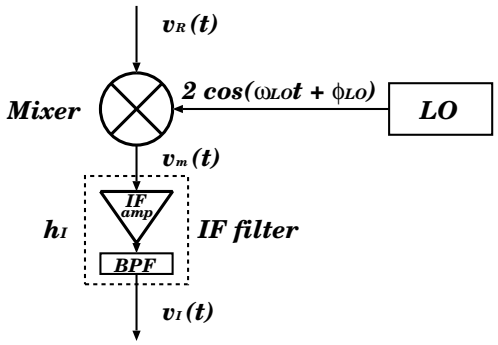

Figure 41: A simple image of a frequency converter.

Let us consider a simple frequency converter, as illustrated in Figure 41, which is composed of a local oscillator (LO), a mixer, and an IF filter. LO generates a reference signal with a frequency ω_{LO} and an initial phase ϕ_{LO}, and the mixer multiplies the RF voltage $v_R(t)$ by the reference signal. Generally speaking, we can arbitrarily choose an amplitude of the LO reference signal. We adopt here $2\cos(\omega_{LO}t + \phi_{LO})$ as the reference signal, since the coefficient 2 gives a simple form of the IF voltage after the down–conversion.

Now, a mixer output voltage $v_m(t)$ has a form:

$$v_m(t) = v_R(t)\, 2\cos(\omega_{LO}t + \phi_{LO}) = v_R(t)\, [e^{i(\omega_{LO}t+\phi_{LO})} + e^{-i(\omega_{LO}t+\phi_{LO})}]. \quad (178)$$

Therefore, in view of the shift theorem given in equation (69), a Fourier transform $\tilde{v}_m(\omega)$ of $v_m(t)$ (i.e., $v_m(t) \Leftrightarrow \tilde{v}_m(\omega)$) is given by

$$\tilde{v}_m(\omega) = \tilde{v}_R(\omega - \omega_{LO})\, e^{i\phi_{LO}} + \tilde{v}_R(\omega + \omega_{LO})\, e^{-i\phi_{LO}}, \tag{179}$$

where $\tilde{v}_R(\omega)$ is again a Fourier transform of the RF voltage $v_R(t)$.

An IF filter is usually composed of a BPF and an IF amplifier. So, let us denote an impulse response and a system function of such an IF filter as $h_I(t)$ and $H_I(\omega)$, respectively. They form, by definition, a Fourier transformation pair:

$$h_I(t) \Leftrightarrow H_I(\omega). \tag{180}$$

Then, an IF voltage signal $v_I(t)$, which is an output of the IF filter, is described through the impulse response $h_I(t)$ as

$$v_I(t) = v_m(t) * h_I(t). \tag{181}$$

Therefore, in view of the convolution theorem given in equation (67), we obtain a Fourier transform $\tilde{v}_I(\omega)$ of the IF voltage signal $v_I(t)$ (i.e., $v_I(t) \Leftrightarrow \tilde{v}_I(\omega)$) in a form:

$$\begin{aligned} \tilde{v}_I(\omega) &= \tilde{v}_m(\omega)\, H_I(\omega) \\ &= [\tilde{v}_R(\omega - \omega_{LO})\, e^{i\phi_{LO}} + \tilde{v}_R(\omega + \omega_{LO})\, e^{-i\phi_{LO}}]\, H_I(\omega). \end{aligned} \tag{182}$$

2.6.2 Cross–Correlation of Fourier Transforms of IF Signals

Let us now consider the frequency conversion in an interferometer using a model schematically illustrated in Figure 42.

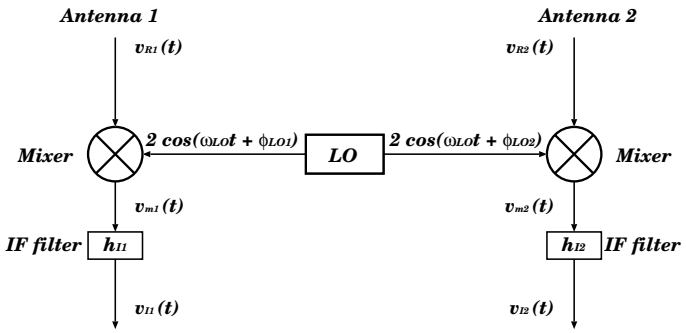

Figure 42: Frequency conversion in an interferometer.

All notations in the previous subsection are used here again with new sufficies 1 and 2, which distinguish voltages and devices in antennas 1 and

2. We assume here that reference signals with a common frequency ω_{LO} are fed to two mixers, based on an image of a connected–element interferometer, which is equipped with a common frequency standard. This assumption is not fully applicable to VLBI, and we will discuss related problems later.

Now, the Fourier transforms of the IF voltages from antennas 1 and 2 are given by

$$\begin{aligned}\tilde{v}_{I1}(\omega) &= [\tilde{v}_{R1}\,(\omega-\omega_{LO})\,e^{i\phi_{LO1}} + \tilde{v}_{R1}\,(\omega+\omega_{LO})\,e^{-i\phi_{LO1}}]\,H_{I1}(\omega),\\ \tilde{v}_{I2}(\omega) &= [\tilde{v}_{R2}\,(\omega-\omega_{LO})\,e^{i\phi_{LO2}} + \tilde{v}_{R2}\,(\omega+\omega_{LO})\,e^{-i\phi_{LO2}}]\,H_{I2}(\omega). \quad (183)\end{aligned}$$

Therefore their cross–correlation is

$$\begin{aligned}\langle \tilde{v}_{I1}(\omega)\,\tilde{v}^*_{I2}(\omega')\rangle &= \langle\ [\tilde{v}_{R1}\,(\omega-\omega_{LO})\,e^{i\phi_{LO1}} + \tilde{v}_{R1}\,(\omega+\omega_{LO})\,e^{-i\phi_{LO1}}]\\ &\quad\times [\tilde{v}^*_{R2}\,(\omega'-\omega_{LO})\,e^{-i\phi_{LO2}} + \tilde{v}^*_{R2}\,(\omega'+\omega_{LO})\,e^{i\phi_{LO2}}]\ \rangle\\ &\quad\times H_{I1}(\omega)\,H^*_{I2}(\omega')\\ &= [\ \langle\tilde{v}_{R1}(\omega-\omega_{LO})\,\tilde{v}^*_{R2}(\omega'-\omega_{LO})\rangle\,e^{i(\phi_{LO1}-\phi_{LO2})}\\ &\quad+\langle\tilde{v}_{R1}(\omega+\omega_{LO})\,\tilde{v}^*_{R2}(\omega'+\omega_{LO})\rangle\,e^{-i(\phi_{LO1}-\phi_{LO2})}\\ &\quad+\langle\tilde{v}_{R1}(\omega-\omega_{LO})\,\tilde{v}^*_{R2}(\omega'+\omega_{LO})\rangle\,e^{i(\phi_{LO1}+\phi_{LO2})}\\ &\quad+\langle\tilde{v}_{R1}(\omega+\omega_{LO})\,\tilde{v}^*_{R2}(\omega'-\omega_{LO})\rangle\,e^{-i(\phi_{LO1}+\phi_{LO2})}\ ]\\ &\quad\times H_{I1}(\omega)\,H^*_{I2}(\omega'). \quad (184)\end{aligned}$$

Since RF voltages $v_{R1}(t)$ and $v_{R2}(t)$ are assumed to be jointly stationary random processes, their Fourier transforms must satisfy the general formula

$$\langle\tilde{v}_{R1}(\omega)\,\tilde{v}^*_{R2}(\omega')\rangle = 2\pi\,S_{v_{R1}v_{R2}}(\omega)\,\delta(\omega-\omega'), \quad (185)$$

as given in equation (84), where $S_{v_{R1}v_{R2}}(\omega)$ is the cross–power spectrum of RF voltages given in equation (177). Therefore, we obtain

$$\begin{aligned}&\langle\tilde{v}_{I1}(\omega)\,\tilde{v}^*_{I2}(\omega')\rangle\\ &= 2\pi[\ S_{v_{R1}v_{R2}}(\omega-\omega_{LO})\,e^{i(\phi_{LO1}-\phi_{LO2})}\,H_{I1}(\omega)\,H^*_{I2}(\omega)\,\delta(\omega-\omega')\\ &\quad+S_{v_{R1}v_{R2}}(\omega+\omega_{LO})\,e^{-i(\phi_{LO1}-\phi_{LO2})}\,H_{I1}(\omega)\,H^*_{I2}(\omega)\,\delta(\omega-\omega')\\ &\quad+S_{v_{R1}v_{R2}}(\omega-\omega_{LO})\,e^{i(\phi_{LO1}+\phi_{LO2})}\,H_{I1}(\omega)\,H^*_{I2}(\omega-2\omega_{LO})\\ &\qquad\qquad\times\ \delta(\omega-\omega'-2\omega_{LO})\\ &\quad+S_{v_{R1}v_{R2}}(\omega+\omega_{LO})\,e^{-i(\phi_{LO1}+\phi_{LO2})}\,H_{I1}(\omega)\,H^*_{I2}(\omega+2\omega_{LO})\\ &\qquad\qquad\times\ \delta(\omega-\omega'+2\omega_{LO})]. \quad (186)\end{aligned}$$

First two terms in RHS of equation (186) are proportional to $\delta(\omega-\omega')$, and therefore, satisfy the condition of jointly stationary random processes, given in equation (84). However, third and fourth terms in RHS of equation (186)

are proportional to $\delta(\omega - \omega' - 2\omega_{LO})$ and $\delta(\omega - \omega' + 2\omega_{LO})$, respectively, and therefore, do not satisfy the condition of equation (84). Consequently, generally speaking, IF voltages after the frequency conversion may not be jointly stationary random processes, i.e. their cross–correlations may depend on time. This is not surprising, because the mixer output $v_m(t)$ in equation (178) is a product of the RF voltage with a LO reference signal, a regular sinusoidal oscillation, which is not random, nor stationary.

2.6.3 Roles of Low–Pass Filters

Nevertheless, if the cross–power spectrum of the RF voltages is band–limited within a certain frequency range, and the IF filters are designed to pass a suitable low–frequency range only, we can make filter out the third and fourth terms, so that resulting IF voltages become jointly stationary.

In fact, if the cross–power spectrum of the RF voltages is band–limited within a range:

$$0 < \omega_1 <\mid \omega \mid < \omega_2, \tag{187}$$

and IF filters $H_{I1}(\omega)$ and $H_{I2}(\omega)$ are both confined within, or proportional to, a rectangular low–pass filter $H_{LP}(\omega)$:

$$H_{LP}(\omega) = \begin{cases} 1 & \text{for } \mid \omega \mid \leq \omega_c, \\ 0 & \text{otherwise,} \end{cases} \tag{188}$$

with a cut–off frequency ω_c, which satisfies conditions:

$$\begin{cases} \omega_2 - \omega_{LO} < \omega_c, \\ \omega_{LO} - \omega_1 < \omega_c, \\ \omega_c < \omega_1 + \omega_{LO}, \end{cases} \tag{189}$$

then, we can eliminate the third and fourth terms in RHS of equation (186), as illustrated in Figure 43.

Figure 43 shows real and imaginary spectral shapes of original complex cross–power spectrum $S_{v_{R1}v_{R2}}(\omega)$ (solid line) in RF–band, and two "shifted" spectra figuring in equation (186), namely, $S_{v_{R1}v_{R2}}(\omega + \omega_{LO})$ (dotted line), and $S_{v_{R1}v_{R2}}(\omega - \omega_{LO})$ (broken line).

$S_{v_{R1}v_{R2}}(\omega)$ is shown here as Hermitian symmetric according to equation (61), since corresponding cross–correlation $R_{v_{R1}v_{R2}}(\tau)$ is real. $S_{v_{R1}v_{R2}}(\omega + \omega_{LO})$ has the same spectral shape as the original $S_{v_{R1}v_{R2}}(\omega)$, but shifted by ω_{LO} towards lower frequency side, while $S_{v_{R1}v_{R2}}(\omega - \omega_{LO})$ also has the same spectral shape, but shifted by ω_{LO} towards higher frequency side.

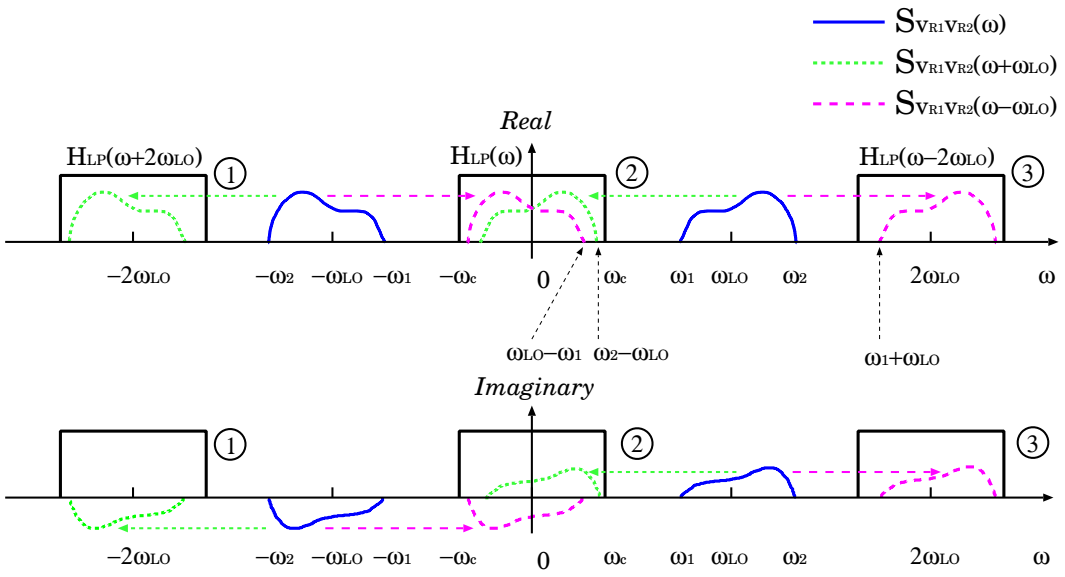

Figure 43: Original complex cross–power spectrum $S_{v_{R1}v_{R2}}(\omega)$ (solid line), and shifted (in frequency) spectra $S_{v_{R1}v_{R2}}(\omega+\omega_{LO})$ (dotted line) and $S_{v_{R1}v_{R2}}(\omega-\omega_{LO})$ (broken line). Boxes assigned with encircled numbers show frequency ranges corresponding to passbands of filters $H_{LP}(\omega+2\omega_{LO})$, $H_{LP}(\omega)$, and $H_{LP}(\omega-2\omega_{LO})$, respectively. Only those parts of spectra which are confined within the central box can be passed through the product of IF filters $H_{I1}(\omega)\,H^*_{I2}(\omega)$ in equation (186).

Both spectral functions $S_{v_{R1}v_{R2}}(\omega+\omega_{LO})$ and $S_{v_{R1}v_{R2}}(\omega-\omega_{LO})$ consist of lower–frequency and higher–frequency parts, corresponding to negative–frequency and positive–frequency parts of the original band–limited RF cross–power spectrum $S_{v_{R1}v_{R2}}(\omega)$. They are confined in three frequency ranges shown by boxes ①, ②, and ③ in Figure 43. These frequency ranges correspond to passbands of rectangular filters $H_{LP}(\omega+2\omega_{LO})$, $H_{LP}(\omega)$, and $H_{LP}(\omega-2\omega_{LO})$, respectively, which are defined by equations (188) and (189).

Let us now examine contributions from 4 terms in RHS of equation (186), one by one.

1. In the first term, the lower–frequency part of $S_{v_{R1}v_{R2}}(\omega-\omega_{LO})$, which is inside the box ②, can be passed through the product of IF filters $H_{I1}(\omega)\,H^*_{I2}(\omega)$, but the higher frequency part, which is inside the box ③, is cut off by the same filter product.

2. In the second term, the higher–frequency part of $S_{v_{R1}v_{R2}}(\omega+\omega_{LO})$, which is inside the box ②, can be passed through the product of IF filters $H_{I1}(\omega)\,H^*_{I2}(\omega)$, but the lower frequency part, which is inside the

box ①, is cut off by the same filter product.

3. In the third term, the entire spectrum $S_{v_{R1}v_{R2}}(\omega-\omega_{LO})$ is cut off by the product of IF filters $H_{I1}(\omega)\,H^*_{I2}(\omega-2\omega_{LO})$, since the lower–frequency part, which is inside the box ②, is cut off by the filter $H^*_{I2}(\omega-2\omega_{LO})$, while the higher–frequency part, which is in ③, is cut off by the filter $H_{I1}(\omega)$.

4. In the fourth term, again, the entire spectrum $S_{v_{R1}v_{R2}}(\omega+\omega_{LO})$ is cut off by the product of IF filters $H_{I1}(\omega)\,H^*_{I2}(\omega+2\omega_{LO})$, since the lower–frequency part, which is inside the box ①, is cut off by the filter $H_{I1}(\omega)$, while the higher–frequency part, which is in ②, is cut off by the filter $H^*_{I2}(\omega+2\omega_{LO})$.

Thus, in the RHS of equation (186), the third and fourth terms are reduced to zero, and only first 2 terms, with the higher–frequency part of $S_{v_{R1}v_{R2}}(\omega+\omega_{LO})$ and lower–frequency part of $S_{v_{R1}v_{R2}}(\omega-\omega_{LO})$, are left.

Therefore, after the low–pass filtering, equation (186) is reduced to

$$\begin{aligned}\langle\tilde{v}_{I1}(\omega)\,\tilde{v}^*_{I2}(\omega')\rangle &\\ = 2\pi[&S_{v_{R1}v_{R2}}(\omega+\omega_{LO})\,e^{-i(\phi_{LO1}-\phi_{LO2})}\\ &+S_{v_{R1}v_{R2}}(\omega-\omega_{LO})\,e^{i(\phi_{LO1}-\phi_{LO2})}]\\ &\times H_{I1}(\omega)\,H^*_{I2}(\omega)\,\delta(\omega-\omega'). \end{aligned}\tag{190}$$

This equation has a proper form of a cross–correlation of Fourier transforms of jointly stationary random processes, which is proportional to $\delta(\omega-\omega')$ as given in equation (84). Consequently, the IF voltages $v_{I1}(t)$ and $v_{I2}(t)$ can now be regarded as jointly stationary random processes.

2.6.4 Relationship between RF Spectrum and IF Spectrum

Since jointly stationary IF voltages must satisfy the general relation:

$$\langle\tilde{v}_{I1}(\omega)\,\tilde{v}^*_{I2}(\omega')\rangle = 2\pi\,S_{v_{I1}v_{I2}}(\omega)\,\delta(\omega-\omega'),\tag{191}$$

where $S_{v_{I1}v_{I2}}(\omega)$ is the cross–power spectrum of IF voltages, equation (190) becomes

$$\begin{aligned}S_{v_{I1}v_{I2}}(\omega) &= [S_{v_{R1}v_{R2}}(\omega+\omega_{LO})\,e^{-i(\phi_{LO1}-\phi_{LO2})}\\ &\quad +S_{v_{R1}v_{R2}}(\omega-\omega_{LO})\,e^{+i(\phi_{LO1}-\phi_{LO2})}]\,H_{I1}(\omega)\,H^*_{I2}(\omega) &(192)\\ &= [S_{v_{R1}v_{R2}}(\omega_{LO}+\omega)\,e^{-i(\phi_{LO1}-\phi_{LO2})}\\ &\quad +S^*_{v_{R1}v_{R2}}(\omega_{LO}-\omega)\,e^{+i(\phi_{LO1}-\phi_{LO2})}]\,H_{I1}(\omega)\,H^*_{I2}(\omega). &(193)\end{aligned}$$

In deriving the last line, we used Hermitian symmetry of cross–power spectrum of real RF voltages v_{R1} and v_{R2}: $S^*_{v_{R1}v_{R2}}(\omega_{LO} - \omega) = S_{v_{R1}v_{R2}}(\omega - \omega_{LO})$. This is a relationship between a cross–power spectrum of IF voltages $S_{v_{I1}v_{I2}}(\omega)$, after the frequency conversion, and a cross–power spectrum of RF voltages $S_{v_{R1}v_{R2}}(\omega)$, before the frequency conversion.

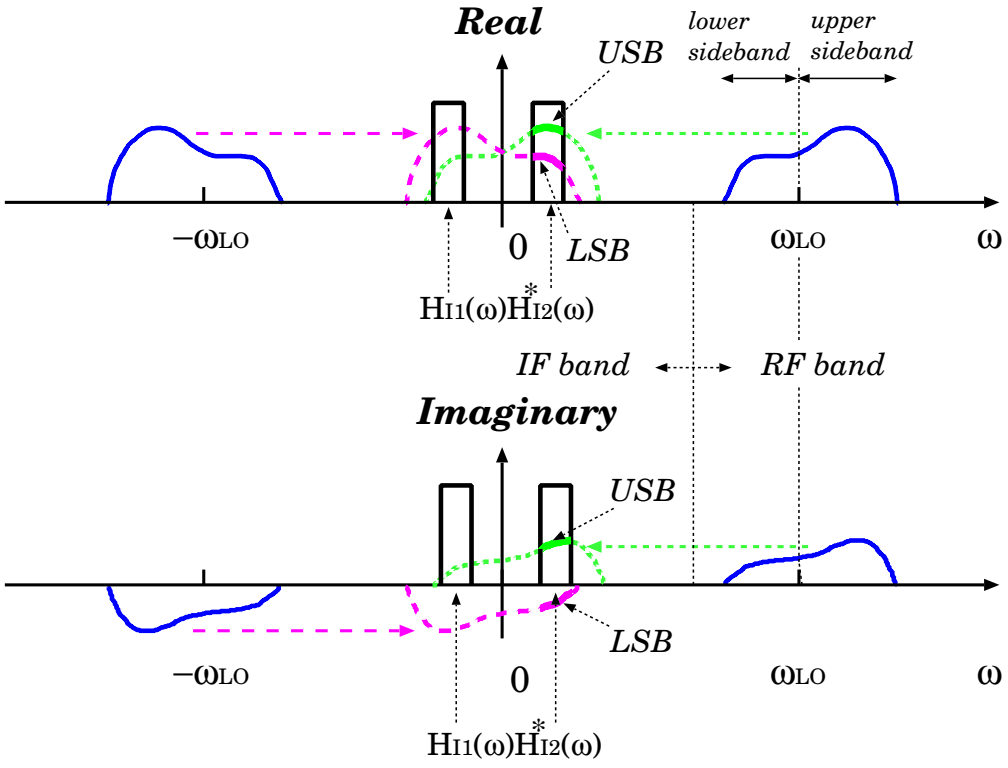

Figure 44: Upper sideband (USB) and lower sideband (LSB) contributions in a cross–power spectrum of IF voltages, which is confined in a passband of the product of IF filters $H_{I1}(\omega)\, H^*_{I2}(\omega)$.

The IF cross–power spectrum is shown in Figure 44, where we assumed the same initial phases of local oscillators, i.e., $\phi_{LO1} - \phi_{LO2} = 0$ for simplicity. In the IF spectrum, the first term in equation (193) shows USB component in the positive frequency side ($\omega \geq 0$) and LSB component in the negative frequency side ($\omega < 0$), while the second term shows LSB component in the positive frequency side ($\omega \geq 0$) and USB component in the negative frequency side ($\omega < 0$).

The IF spectrum is Hermitian symmetric, as expected from the general property of a cross–power spectrum of a real cross–correlation shown in

equation (61). Since the cross–correlation can be described by the positive–frequency region of the Hermitian symmetric cross–power spectrum, as we saw in equation (63), we will pay our attension to the positive frequency region ($\omega \geq 0$) of the IF spectrum only. Then, the original RF cross–power spectrum, shown by solid lines in Figure 44, is down–converted to IF band, and the upper sideband (USB) of $S_{v_{R1}v_{R2}}(\omega + \omega_{LO})$ (shown by a broken line), and the lower sideband (LSB) of $S_{v_{R1}v_{R2}}(\omega - \omega_{LO})$ (shown by a dotted line) are passed through the product of the IF filters $H_{I1}(\omega)\, H^*_{I2}(\omega)$ in the positive IF frequency region, as we see in Figure 44.

Thus, we can notice here following properties of the IF spectrum:

- Spectral shape of the RF spectrum is preserved, and just shifted into the IF band.
- Upper sideband (USB) and lower sideband (LSB) components are superposed in the IF passband of $H_{I1}(\omega)\, H^*_{I2}(\omega)$.
- In the positive frequency range of the IF spectrum, $\omega \geq 0$, spectrum of the LSB component is reversed, compared with the positive frequency part of the original RF spectrum.

2.7 White Fringe in IF Band

Before going to the next step, i.e. to the "delay tracking" and "fringe stopping", we will make sure if we still need this step after the frequency conversion. For this purpose, we consider what will come out if we directly cross–correlate IF voltages (Figure 45).

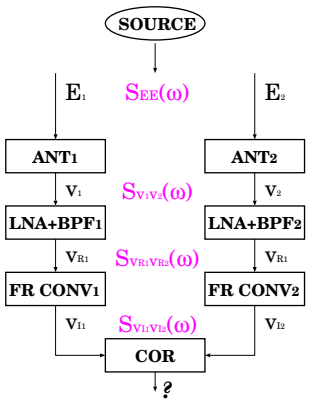

Figure 45: What will come out if we cross–correlate IF voltages?

Specifically, we will combine the cross–power spectra obtained above, and express the cross–power spectrum of the IF voltages in terms of the radio source information. Then, we will calculate a cross–correlation of the IF voltages by means of the inverse Fourier transformation.

2.7.1 Cross–Power spectra of Received, RF and IF Voltages

So far, we derived

1. the cross–power spectrum of received voltages:

$$S_{v_1 v_2}(\omega) = \frac{1}{4} A_0 \, e^{-i\omega\tau_{g_0}} \, \mathcal{V}(\omega),$$

given in equation (168), where A_0 is the geometrical mean of effective apertures of antennas, and $\mathcal{V}(\omega)$ is the complex visibility:

$$\mathcal{V}(\omega) = \oint A_N(\boldsymbol{\sigma}) \, I_\nu(\boldsymbol{\sigma}) \, e^{-i2\pi \boldsymbol{D}_\lambda \cdot \boldsymbol{\sigma}} \, d\Omega,$$

given in equation (165), where $A_N(\boldsymbol{\sigma})$ is the normalized power pattern of an interferometer, $\boldsymbol{\sigma} = \boldsymbol{s} - \boldsymbol{s}_0$ is an offset vector showing a certain direction $\boldsymbol{s}$ in a source with respect to the reference direction $\boldsymbol{s}_0$, $I_\nu(\boldsymbol{\sigma})$ is the source intensity distribution, and $\boldsymbol{D}_\lambda = \boldsymbol{D}/\lambda$ is a baseline vector normalized by a wave length λ,

2. the cross–power spectrum of RF voltages after the RF filters (amplifiers and BPFs) with system functions $H_{R1}(\omega)$ and $H_{R2}(\omega)$:

$$S_{v_{R1} v_{R2}}(\omega) = S_{v_1 v_2}(\omega) \, H_{R1}(\omega) \, H^*_{R2}(\omega),$$

given in equation (177),

3. the cross–power spectrum of IF voltages after the frequency conversion:

$$\begin{aligned} S_{v_{I1} v_{I2}}(\omega) = [&S_{v_{R1} v_{R2}}(\omega_{LO} + \omega) \, e^{-i(\phi_{LO1} - \phi_{LO2})} \\ &+ S^*_{v_{R1} v_{R2}}(\omega_{LO} - \omega) \, e^{i(\phi_{LO1} - \phi_{LO2})}] \\ &\times H_{I1}(\omega) \, H^*_{I2}(\omega), \end{aligned}$$

given in equation (193), where ω_{LO} is the LO reference frequency, ϕ_{LO1} and ϕ_{LO2} are initial phases of local oscillators, and $H_{I1}(\omega)$ and $H_{I2}(\omega)$ are system functions of the IF filters.

Combining last two equations, i.e., equations (177) and (193), we have

$$\begin{aligned} &S_{v_{I1} v_{I2}}(\omega) \\ &= [H_{R1}(\omega_{LO} + \omega) \, H^*_{R2}(\omega_{LO} + \omega) \, S_{v_1 v_2}(\omega_{LO} + \omega) \, e^{-i(\phi_{LO1} - \phi_{LO2})} \\ &\quad + H^*_{R1}(\omega_{LO} - \omega) \, H_{R2}(\omega_{LO} - \omega) \, S^*_{v_1 v_2}(\omega_{LO} - \omega) \, e^{i(\phi_{LO1} - \phi_{LO2})}] \\ &\qquad \times H_{I1}(\omega) \, H^*_{I2}(\omega). \end{aligned} \tag{194}$$

2.7.2 Combined RF and IF Filters

Since system functions of real impulse responses of IF filters must satisfy Hermitian symmetry as shown in equation (72), we have

$$H_{I1}(\omega) = H_{I1}^*(-\omega), \quad \text{and} \quad H_{I2}^*(\omega) = H_{I2}(-\omega).$$

Then, introducing new system functions $H_1(\omega)$ and $H_2(\omega)$ for responses of combined RF and IF filters:

$$\begin{aligned} H_1(\omega) &= H_{R1}(\omega_{LO} + \omega)\, H_{I1}(\omega), \\ H_2(\omega) &= H_{R2}(\omega_{LO} + \omega)\, H_{I2}(\omega), \end{aligned} \tag{195}$$

we can transform equation (194) into

$$\begin{aligned} &S_{v_{I1}v_{I2}}(\omega) \\ &= H_1(\omega)\, H_2^*(\omega)\, S_{v_1v_2}(\omega_{LO} + \omega)\, e^{-i(\phi_{LO1} - \phi_{LO2})} \\ &\quad + H_1^*(-\omega)\, H_2(-\omega)\, S_{v_1v_2}^*(\omega_{LO} - \omega)\, e^{i(\phi_{LO1} - \phi_{LO2})}. \end{aligned} \tag{196}$$

Hereafter, we call the complex function $H_1(\omega)\, H_2^*(\omega)$ "bandpass characteristics of combined RF and IF filters".

2.7.3 Cross–Power Spectra of IF Voltages in Terms of Complex Visibility

Now, expressing $S_{v_1v_2}(\omega)$ through the complex visibility on the basis of equation (168), we obtain

$$\begin{aligned} &S_{v_{I1}v_{I2}}(\omega) \\ &= \frac{1}{4}\, A_0 \,\{e^{-i[(\omega_{LO}+\omega)\tau_{g_0} + \phi_{LO1} - \phi_{LO2}]}\, \mathcal{V}(\omega_{LO} + \omega)\, H_1(\omega)\, H_2^*(\omega) \\ &\qquad + e^{i[(\omega_{LO}-\omega)\tau_{g_0} + \phi_{LO1} - \phi_{LO2}]}\, \mathcal{V}^*(\omega_{LO} - \omega)\, H_1^*(-\omega)\, H_2(-\omega)\}, \end{aligned} \tag{197}$$

which describes the cross–power spectrum of the IF voltages $S_{v_{I1}v_{I2}}(\omega)$ in terms of source intensity distribution, geometry of observation, filter responses, and parameters of the frequency conversion. Of course, this equation satisfies the Hermitian symmetry condition:

$$S_{v_{I1}v_{I2}}(\omega) = S_{v_{I1}v_{I2}}^*(-\omega),$$

required for real processes $v_{I1}(t)$ and $v_{I2}(t)$.

Representing the complex visibility in terms of visibility amplitude and visibility phase, as introduced in equation (166), we have

$$\begin{aligned} \mathcal{V}(\omega_{LO}+\omega) &= \mid \mathcal{V}(\omega_{LO}+\omega) \mid \, e^{i\Phi_v(\omega_{LO}+\omega)}, \\ \mathcal{V}^*(\omega_{LO}-\omega) &= \mid \mathcal{V}(\omega_{LO}-\omega) \mid \, e^{-i\Phi_v(\omega_{LO}-\omega)}. \end{aligned} \tag{198}$$

Also we represent the bandpass characteristics of combined RF and IF filters in terms of its amplitude $\mid H_1(\omega)\, H_2^*(\omega) \mid$ and phase $\Phi_b(\omega)$:

$$\begin{aligned} H_1(\omega)\, H_2^*(\omega) &= \mid H_1(\omega)\, H_2^*(\omega) \mid \, e^{i\Phi_b(\omega)}, \\ H_1^*(-\omega)\, H_2(-\omega) &= \mid H_1(-\omega)\, H_2^*(-\omega) \mid e^{-i\Phi_b(-\omega)}. \end{aligned} \tag{199}$$

Then, the cross–power spectrum of IF signals is now reduced to

$$\begin{aligned} & S_{v_{I1}v_{I2}}(\omega) \\ &= \frac{1}{4}\, A_0 \, \mid \mathcal{V}(\omega_{LO}+\omega) \mid\mid H_1(\omega)\, H_2^*(\omega) \mid \, e^{-i[(\omega_{LO}+\omega)\tau_{g_0}+\phi_{LO1}-\phi_{LO2}-\Phi_v(\omega_{LO}+\omega)-\Phi_b(\omega)]} \\ &+\frac{1}{4}\, A_0 \, \mid \mathcal{V}(\omega_{LO}-\omega) \mid\mid H_1(-\omega)\, H_2^*(-\omega) \mid \, e^{i[(\omega_{LO}-\omega)\tau_{g_0}+\phi_{LO1}-\phi_{LO2}-\Phi_v(\omega_{LO}-\omega)-\Phi_b(-\omega)]}. \end{aligned} \tag{200}$$

If we denote amplitudes and phases of the first and second terms of the RHS of equation (200) as $A_p(\omega)$, $\Phi_p(\omega, t)$ and $A_n(\omega)$, $\Phi_n(\omega, t)$, respectively, we have

$$S_{v_{I1}v_{I2}}(\omega) = A_p(\omega)\, e^{-i\Phi_p(\omega,\, t)} + A_n(\omega)\, e^{-i\Phi_n(\omega,\, t)}, \tag{201}$$

with

$$\begin{aligned} A_p(\omega) &= \frac{1}{4}\, A_0 \, \mid \mathcal{V}(\omega_{LO}+\omega) \mid\mid H_1(\omega)\, H_2^*(\omega) \mid, \\ A_n(\omega) &= \frac{1}{4}\, A_0 \, \mid \mathcal{V}(\omega_{LO}-\omega) \mid\mid H_1(-\omega)\, H_2^*(-\omega) \mid, \end{aligned} \tag{202}$$

and

$$\begin{aligned} \Phi_p(\omega,\, t) &= (\omega_{LO}+\omega)\tau_{g_0} + \phi_{LO1} - \phi_{LO2} - \Phi_v(\omega_{LO}+\omega) - \Phi_b(\omega), \\ \Phi_n(\omega,\, t) &= -(\omega_{LO}-\omega)\tau_{g_0} - \phi_{LO1} + \phi_{LO2} + \Phi_v(\omega_{LO}-\omega) + \Phi_b(-\omega). \end{aligned} \tag{203}$$

Here, we describe phases as functions of frequency and time, taking into account time variability of the geometric delay τ_{g_0}. Equations (201), (202), and (203) show amplitude and phase spectra of the cross–power spectrum of IF signals, which can be contrasted to those given in equations (101), (102), and (103), for the simple interferometer model. Again, first term in

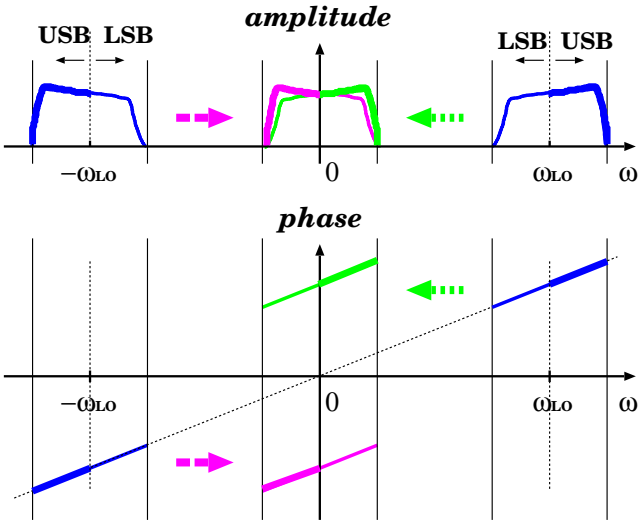

Figure 46: Amplitudes and phases of two terms in a cross–power spectrum of IF voltages. Solid lines show original RF spectrum, and dotted and broken lines correspond to first and second terms of the RHS of equation (200), respectively. For simplicity, only contribution of geometric delay is shown in phase spectrum.

the RHS of equation (200) shows USB component in the positive frequency range ($\omega \geq 0$) and LSB component in the negative frequency range ($\omega < 0$), while second term shows LSB component in the positive frequency range ($\omega \geq 0$) and USB component in the negative frequency range ($\omega < 0$).

Figure 46 shows the amplitude and phase spectra of two terms of the IF cross–power spectrum in the RHS of equation (200), which can be compared with Figure 21 for the simple interferometer. For simplicity, we showed only contribution of geometric delay τ_{g0} in this figure, ignoring small effects of ϕ_{LO1}, ϕ_{LO2}, Φ_v, and Φ_b terms. The IF band here is assumed to be a "video–band" (or "baseband"), which includes zero frequency, or "DC (direct current)", component at $\omega = 0$.

2.7.4 "Expected Correlation" of Voltages

Now, we introduce a new term: "expected correlation".

What we call **"expected correlation"** is a statistical expectation of a product of jointly stationary random voltages at a certain instance of time, or a cross–correlation of the voltages at time difference $\tau = 0$. For example, expected correlation of IF voltages is $\langle v_{I1}(t)\, v_{I2}(t) \rangle = R_{v_{I1} v_{I2}}(0)$.

This is not an actual correlation, which always requires some non–zero integration time. This is not a mere product of voltages, but a statistical

expectation of the product. In particular, this contains **signals** only, but **not noise** contribution, as far as noises are uncorrelated.

If we had an idealized correlator, which would be capable of multiplying two IF voltages, and immediately calculating a statistical expectation of the product, such a correlator would yield the expected correlation of the IF voltages $R_{v_{I1}v_{I2}}(0)$.

According to equation (197), we can calculate the expected correlation of the IF voltages by an inverse Fourier transformation of the cross–power spectrum $S_{v_{I1}v_{I2}}(\omega)$, with $\tau = 0$:

$$R_{v_{I1}v_{I2}}(0) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} S_{v_{I1}v_{I2}}(\omega)\, d\omega = \frac{1}{\pi}\, \Re \int\limits_{0}^{\infty} S_{v_{I1}v_{I2}}(\omega)\, d\omega$$

$$= \frac{A_0}{4\pi}\, \Re \left[e^{-i(\omega_{LO}\tau_{g_0} + \phi_{LO1} - \phi_{LO2})} \int\limits_{0}^{\infty} \mathcal{V}(\omega_{LO} + \omega)\, e^{-i\omega\tau_{g_0}}\, H_1(\omega)\, H_2^*(\omega)\, d\omega \right.$$

$$\left. + e^{i(\omega_{LO}\tau_{g_0} + \phi_{LO1} - \phi_{LO2})} \int\limits_{0}^{\infty} \mathcal{V}^*(\omega_{LO} - \omega)\, e^{-i\omega\tau_{g_0}}\, H_1^*(-\omega)\, H_2(-\omega)\, d\omega \right], \quad (204)$$

where we used the property of a Hermitian symmetric cross–power spectrum given in equation (72), in order to restrict range of integration to the positive frequency side only.

2.7.5 White Fringe in Expected Correlation of IF Signals

As an example, let us calculate the expected correlation of IF voltages in a simple case of real, even, and rectangular combined RF and IF filters and constant (white noise) visibility of a source, as we did in the simple interferometer model of Section 1.3. Let a gain factor, bandwidth, and center frequency of the combined RF and IF filters be G, $\Delta\omega = 2\pi B$, and ω_I, correspondingly, as illustrated in Figure 47, i.e.,

$$H_1(\omega)\, H_2^*(\omega) = H_1^*(-\omega)\, H_2(-\omega) = \begin{cases} G & \text{if} \quad \omega_I - \frac{\Delta\omega}{2} \leq \mid \omega \mid \leq \omega_I + \frac{\Delta\omega}{2}, \\ 0 & \text{otherwise,} \end{cases} \quad (205)$$

and the visibility be

$$\mathcal{V}(\omega) \cong \mathcal{V} \quad \text{for} \quad \omega_I - \frac{\Delta\omega}{2} \leq \mid \omega \mid \leq \omega_I + \frac{\Delta\omega}{2}. \quad (206)$$

Then, we obtain from equation (204)

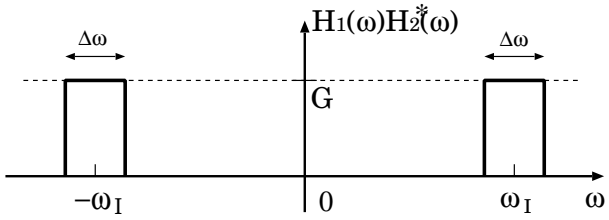

Figure 47: Rectangular response of a real filter.

$$R_{v_{I1}v_{I2}}(0) = \frac{A_0\,G}{4\pi}\,\Re\left[e^{-i(\omega_{LO}\tau_{g_0}+\phi_{LO1}-\phi_{LO2})}\,\mathcal{V}\int_{\omega_I-\frac{\Delta\omega}{2}}^{\omega_I+\frac{\Delta\omega}{2}} e^{-i\omega\tau_{g_0}}\,d\omega\right.$$
$$\left. + e^{i(\omega_{LO}\tau_{g_0}+\phi_{LO1}-\phi_{LO2})}\,\mathcal{V}^*\int_{\omega_I-\frac{\Delta\omega}{2}}^{\omega_I+\frac{\Delta\omega}{2}} e^{-i\omega\tau_{g_0}}\,d\omega\right]. \quad (207)$$

Since the integral in equation (207) yields the sinc function, as we saw in equation (104):

$$\frac{1}{\pi}\int_{\omega_I-\frac{\Delta\omega}{2}}^{\omega_I+\frac{\Delta\omega}{2}} e^{-i\omega\tau_{g_0}}\,d\omega = 2Be^{-i\omega_I\tau_{g_0}}\frac{\sin(\pi B\tau_{g_0})}{\pi B\tau_{g_0}},$$

and the complex visibility is represented in terms of its amplitude and phase:

$$\mathcal{V} =\mid \mathcal{V} \mid e^{i\Phi_v},$$

we can further reduce equation (207) to

$$\begin{aligned} R_{v_{I1}v_{I2}}(0) = {} & \frac{A_0GB \mid \mathcal{V} \mid}{2}\frac{\sin(\pi B\tau_{g_0})}{\pi B\tau_{g_0}} \\ \times {} & [\cos(\omega_{LO}\tau_{g_0}+\omega_I\tau_{g_0}+\phi_{LO1}-\phi_{LO2}-\Phi_v) \\ + {} & \cos(\omega_{LO}\tau_{g_0}-\omega_I\tau_{g_0}+\phi_{LO1}-\phi_{LO2}-\Phi_v)], \end{aligned} \quad (208)$$

where first and second cosine terms show USB and LSB contributions, respectively.

Earlier, we derived the white fringe in equation (104) for the "RF correlation" case, based on the "simple interferometer" model:

$$R_{v_1v_2}(0) = 2\,A_0\,G\,B\,S(\omega_0)\,\frac{\sin(\pi B\tau_g)}{\pi B\tau_g}\,\cos(\omega_0\tau_g),$$

where ω_0 is an RF band center frequency, and $S(\omega_0)$ is the amplitude of the white–noise power spectrum of a single–polarization component of an incident electric field from a point source (Figure 48).

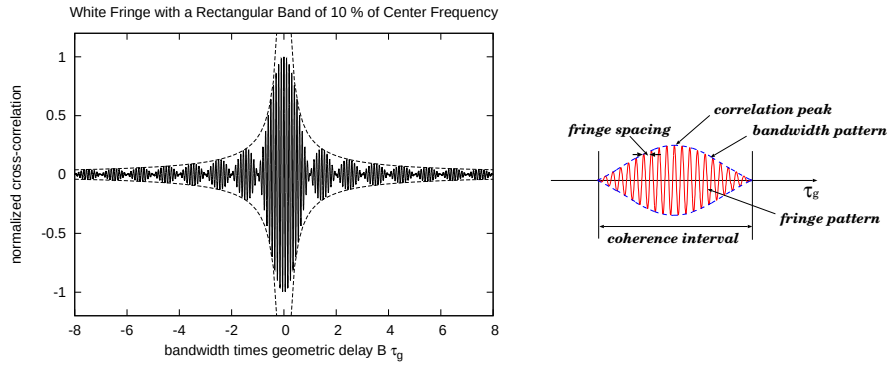

Figure 48: White fringe.

Now, for a more realistic interferometer with the frequency conversion, we derived in equation (208) the expected correlation of the IF voltages:

$$R_{v_{I1}v_{I2}}(0) = \frac{A_0 GB \mid \mathcal{V} \mid}{2} \frac{\sin(\pi B\tau_{g_0})}{\pi B\tau_{g_0}}$$
$$\times \ [\cos(\omega_U\tau_{g_0} + \phi_{LO1} - \phi_{LO2} - \Phi_v) + \cos(\omega_L\tau_{g_0} + \phi_{LO1} - \phi_{LO2} - \Phi_v)],$$

where we introduced RF band center frequencies $\omega_U = \omega_{LO} + \omega_I$ and $\omega_L = \omega_{LO} - \omega_I$, for the USB and LSB cases, respectively. Each of the USB and LSB terms shows the white fringe form, i.e. the fringe pattern enclosed by the bandwidth pattern, very similar to the one in the "RF correlation" case illustrated in Figure 48. Note that $S(\omega_0) =\mid \mathcal{V} \mid /4$ in the present notation.

Thus the white fringe nature of the expected correlation remains unaltered irrespective of whether we take the correlation in RF or IF bands, except for the emergence and superposition of USB and LSB contributions in the IF correlation case. In particular, the very narrow fringe spacing and the rapid fringe oscillation due to time variation of the geometric delay τ_{g_0} in the fringe pattern are governed by RF frequency, as $\omega_U\tau_{g_0}$ and $\omega_L\tau_{g_0}$ terms in the cosine fringe patterns show, though we obtained the expected correlation in IF band.

This again implies that we cannot detect the white fringe of a continuum spectrum source except in a very narrow coherence interval, roughly determined by $\Delta\tau_B = 2/B$ around $\tau_{g_0} = 0$. Also, the rapid fringe oscillation of the expected correlation (which we called the "correlator output" in subsection

1.3.9) makes it almost impossible to integrate (time–average) the product of voltages for a meaningful duration of time. Therefore, we definitely need suitable means to overcome these difficulties.

2.8 Delay Tracking and Fringe Stopping

We learned in subsections 1.3 and 2.7 that modern interferometers must be equipped with special means, for compensating the delay and stopping the fringe oscillation of the expected correlation. Tables 1 and 2 illustrated how interferometric observations would be difficult if such special means do not exist. These means are the "delay tracking" and "fringe stopping". They are indispensable processing units in a realistic interferometer (Figure 49).

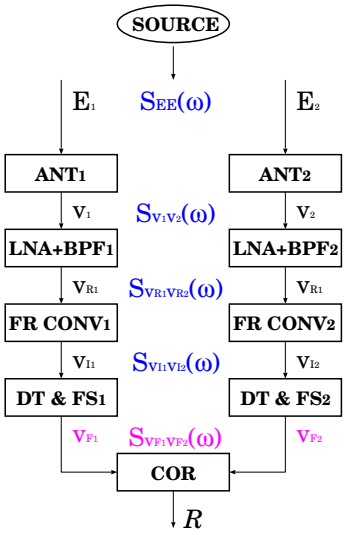

Figure 49: How does cross–power spectrum of IF voltages look like after delay tracking and fringe stopping?

Meanwhile, one can understand the delay tracking and fringe stopping as "tracking of a radio source by an interferometer".

You could wonder why we still need the tracking despite that we are already tracking the source by element antennas of the interferometer. Of course, element antennas must track the source in order to catch and keep it within their beams. Otherwise, the antennas simply could not receive any radio wave from the source.

However, discussions in subsections 1.3 and 2.7 showed that even though the antennas properly track the source and provide us received voltages from the source, we would get almost nothing if we just directly feed the voltages

to the correlator. This is because source signals contained in the voltages then most likely have a finite delay which is out of the very narrow "field of view of the interferometer", i.e. the coherence interval around the zero delay.

Moreover, even when the source happens to be within the coherence interval, the expected correlation oscillates very rapidly due to the source motion across the very dense fringe pattern, making meaningful integration almost impossible.

Thus, in order to get scientific information, we need to "steer" the field of view of the interferometer, i.e. the coherence interval and the fringe pattern contained therein, following the moving source, so that we can stably catch the source within the coherence interval and keep it stopped with respect to the fringe pattern. This is quite similar what a single–dish radio telescope does with its antenna beam for tracking a source. Thus, this is the "additional tracking" of a source "by an interferometer".

Now we would like to discuss how we can realize such tracking in a relaistic interferometer.

2.8.1 Radio Source Tracking by an Interferometer

What is needed is already clear from previous discussions. If we are always able to feed the source signals received by element antennas to a correlator **with no delay**, then we can constantly find the source at the very center of the coherence interval, or at the correlation peak of the white fringe. Otherwise speaking, we need to manage received signals so that signals corresponding to the same wave front arrive at the correlator at the same time.

An obvious way to achieve such a goal is to place element antennas, with equal–length cables to a correlator, onto a rigid plate and drive the plate so

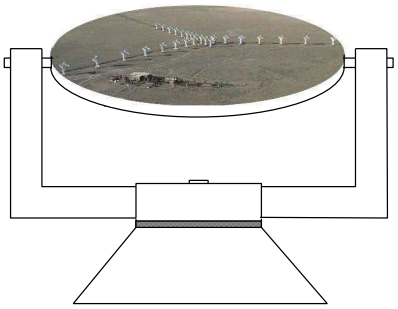

Figure 50: An interferometer on a "steerable plate" would automatically perform the delay tracking and fringe stopping (photo courtesy of NRAO/AUI/NSF).

that its normal always follows an observed source as illustrated in Figure 50.

Although conceptually simple, making such a steerable plate is technologically quite difficult even for a small–scale connected element interferometer array and totally inconceivable for a VLBI.

There is another way to achieve the goal. Suppose that we are able to shift one of antennas of a single–baseline interferometer along a ray path from an astronomical radio source to a point where geometric delay between the two antennas is equal to zero. Then again we can feed the signals with no delay to the correlator. In such a scheme, the moving antenna would trace a surface of a sphere with a diameter equal to the baseline length as the source changes its direction with respect to the baseline (Figure 51).

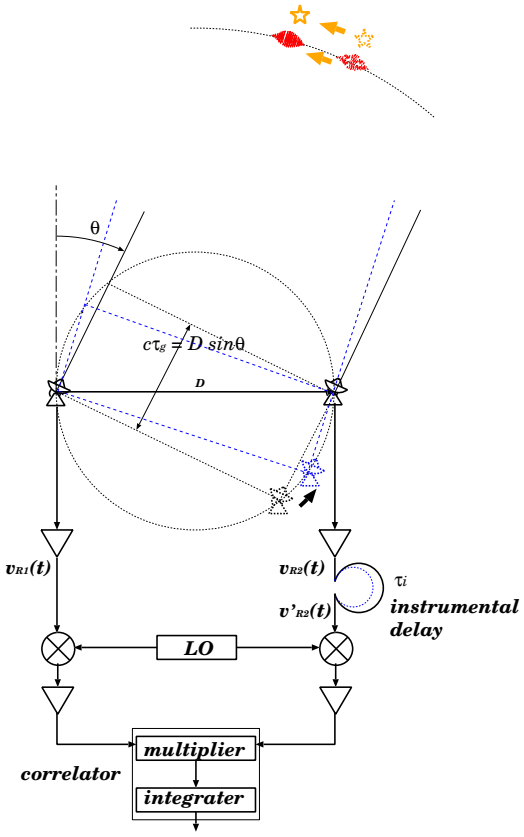

Figure 51: Insertion of an instrumental delay τ_i in the RF band, which compensates the geometric delay τ_g, is equivalent to shifting an antenna along a ray path by $c\tau_g$.

Though it is also technologically difficult to move an antenna along the ray path, we have an equivalent way which can instrumentally realize the same effect.

Indeed, if we insert an instrumentally implemented time variable delay

τ_i which is always equal to the geometric delay τ_{g_0} into the RF–band transmission system of one of the antennas as shown in Figure 51, then we can feed the received signals to the correlator with no delay. We call such an instrumentally implemented time variable delay as "**instrumental delay**".

The above statement must be intuitively obvious. However, it would be still useful to logically trace the effect of the delay insertion at the RF band.

2.8.2 Insertion of Instrumental Delay at RF Band

Insertion of an instrumental delay τ_i, in the RF signal transmission system of antenna 2, would make the RF voltage signal $v'_{R2}(t)$, after the delay circuit, and its Fourier transform $\tilde{v}'_{R2}(\omega)$, to be equal to

$$v'_{R2}(t) = v_{R2}(t-\tau_i), \quad \text{and therefore,} \quad \tilde{v}'_{R2}(\omega) = \tilde{v}_{R2}(\omega)\, e^{-i\omega\tau_i}, \tag{209}$$

where $v_{R2}(t)$, and $\tilde{v}_{R2}(\omega)$, correspond to the RF voltage signal before the instrumental delay. We used the shift theorem given in equation (69), in deriving equation (209).

Therefore, taking a cross–correlation of Fourier transforms $\tilde{v}_{R1}(\omega)$ and $\tilde{v}'_{R2}(\omega)$ of the RF voltage signals, we see

$$\langle \tilde{v}_{R1}(\omega)\, \tilde{v}'^{*}_{R2}(\omega') \rangle = \langle \tilde{v}_{R1}(\omega)\, \tilde{v}^{*}_{R2}(\omega') \rangle\, e^{i\omega'\tau_i}. \tag{210}$$

Since we assumed that the RF signals $v_{R1}(t)$ and $v_{R2}(t)$ are jointly stationary random processes, $v_{R1}(t)$ and $v'_{R2}(t)$ must also be jointly stationary random processes. Hence, the cross–correlations of their Fourier transforms and cross–power spectra must be related to each other as:

$$\begin{aligned} \langle \tilde{v}_{R1}(\omega)\, \tilde{v}'^{*}_{R2}(\omega') \rangle &= 2\pi S_{v_{R1}v'_{R2}}(\omega)\, \delta(\omega-\omega'), \\ \langle \tilde{v}_{R1}(\omega)\, \tilde{v}^{*}_{R2}(\omega') \rangle &= 2\pi S_{v_{R1}v_{R2}}(\omega)\, \delta(\omega-\omega'), \end{aligned} \tag{211}$$

in view of equation (84). Comparing equations (210) and (211), we obtain a relation between the cross–power spectrum $S_{v_{R1}v'_{R2}}(\omega)$ of the RF voltage signal $v_{R1}(t)$ of antenna 1 and the delay–inserted RF voltage signal $v'_{R2}(t)$ of antenna 2, and the cross–power spectrum $S_{v_{R1}v_{R2}}(\omega)$ of the RF voltages without delay insertion:

$$S_{v_{R1}v'_{R2}}(\omega) = S_{v_{R1}v_{R2}}(\omega)\, e^{i\omega\tau_i}, \tag{212}$$

which, in view of equations (168) and (177), leads to

$$\begin{aligned} S_{v_{R1}v'_{R2}}(\omega) &= S_{v_1v_2}(\omega)\, H_{R1}(\omega)\, H^{*}_{R2}(\omega)\, e^{i\omega\tau_i} \\ &= \frac{1}{4}\, A_0\, e^{-i\omega\Delta\tau_g}\, \mathcal{V}(\omega)\, H_{R1}(\omega)\, H^{*}_{R2}(\omega), \end{aligned} \tag{213}$$

where $H_{R1}(\omega)$ and $H_{R2}(\omega)$ are the system functions of the RF filters, $\mathcal{V}(\omega)$ is the complex visibility of the source, A_0 is the geometric mean of effective apertures of two antennas, and $\Delta\tau_g$ is a difference between the geometric delay and the instrumental delay:

$$\Delta\tau_g = \tau_{g_0} - \tau_i,$$

which we call hereafter "**residual delay**".

As we see, an only different point between combined equations (168) and (177) and this equation (213) is that the geometric delay τ_{g_0} is replaced by the residual delay $\Delta\tau_g = \tau_{g_0} - \tau_i$.

Now we can easily derive an expected correlation $R_{v_{I1}v'_{I2}}(0)$ of IF votages after the freqency conversion, where $v_{I1}(t)$ is an IF voltage of antenna 1 and $v'_{I2}(t)$ is an IF voltage with delay inserted at RF band of antenna 2, in the same way as we derived equation (208), assuming again a rectangular filter, given in equation (205), and constant visibility over receiving bandwidth, given in equation (206). The result is

$$\begin{aligned} R_{v_{I1}v'_{I2}}(0) = & \frac{A_0 GB \mid \mathcal{V} \mid}{2} \frac{\sin(\pi B\Delta\tau_g)}{\pi B\Delta\tau_g} \\ & \times \quad [\cos(\omega_{LO}\Delta\tau_g + \omega_I\Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v) \\ & + \quad \cos(\omega_{LO}\Delta\tau_g - \omega_I\Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v)], \end{aligned} \tag{214}$$

where notations are the same with those in equation (208).

Thus, as long as $\tau_i = \tau_{g_0}$, no delay effect is left in the RF cross–power spectrum, given in equation (213), and, therefore, also in the expected correlation of IF voltages in equation (214). This means that the insertion of an instrumental delay, exactly equal to the geometric delay, at the RF band could effectively reduce τ_g in equation (104), or τ_{g_0} in equation (208), to zero. This would allow us to find the white fringe always at the center of the coherence interval, and to integrate the product of voltages, as long as we wish. Or, otherwise speaking, we could realize both delay tracking and fringe stopping by a single operation, namely by the insertion of a time variable instrumental delay, if we were allowed to do it at the RF band.

Unfortunately, it is quite difficult, at least in the current level of technology, to realize any delay insertion at RF frequencies, higher than several GHz. Therefore, the delay insertion is usually performed at the IF band.

2.8.3 Requirements to the Delay Tracking and Fringe Stopping

Before proceeding with discussions on actual realization of the delay tracking and fringe stopping at the IF band, let us here estimate the ranges of variations, and accuracies required to theoretical predictions, of the geometric

delay τ_{g_0} and its time rate $\dot{\tau}_{g_0}$. For simplicity, we will implicitly rely upon an image of the delay insertion at the RF band discussed above, since this does not cause any loss of generality in present discussions.

If we denote the baseline vector and the unit vector directed to the radio source as $\boldsymbol{D}$ and $\boldsymbol{s}_0$, respectively, the geometric delay τ_{g_0} and its time rate (delay rate) $\dot{\tau}_{g_0}$, due to the diurnal rotation of the Earth, are given by:

$$\tau_{g_0} = \frac{\boldsymbol{D} \cdot \boldsymbol{s}_0}{c}, \tag{215}$$

$$\dot{\tau}_{g_0} = \frac{\dot{\boldsymbol{D}} \cdot \boldsymbol{s}_0}{c} = \frac{(\boldsymbol{\omega} \times \boldsymbol{D}) \cdot \boldsymbol{s}_0}{c}, \tag{216}$$

where $\boldsymbol{\omega}$ is the angular velocity vector of the rotation of the Earth ($|\boldsymbol{\omega}| \simeq 7.3 \times 10^{-5}$ rad/sec) and $c \simeq 3 \times 10^8$ m/sec is the light velocity. Of course, equation (216) is valid for ground–based interferometers only, and not for space–VLBI baselines.

For a connected–element radio interferometer (CERI) of 300 m baseline, say, the maximum geometric delay:

$$\tau_{g_{max}} = \frac{D}{c} = 1 \ \mu\text{sec},$$

is obtained when the source direction is parallel to the baseline, and the maximum delay rate:

$$\dot{\tau}_{g_{max}} = \frac{\omega D}{c} = 70 \ \text{psec/sec},$$

is obtained when $\boldsymbol{\omega} \perp \boldsymbol{D}$ and $(\boldsymbol{\omega} \times \boldsymbol{D}) \parallel \boldsymbol{s_0}$.

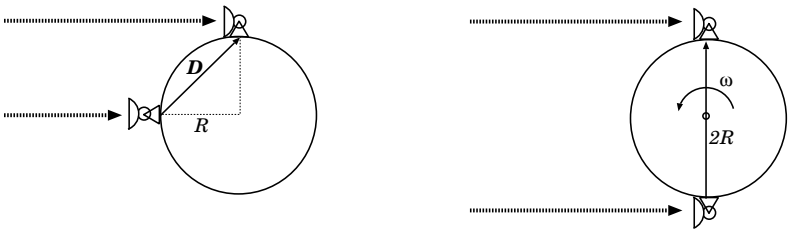

Figure 52: Configurations corresponding to the maximum geometric delay (left) and the maximum delay rate (right) in the ground–based VLBI.

For the ground–based VLBI, the maximum geometric delay and the maximum delay rate are obtained in cases shown in Figure 52. If we denote the

Earth's radius as $R_\oplus = 6300$ km, the maximum values are

$$\tau_{g_{max}} = \frac{R_\oplus}{c} = 21 \text{ msec}, \tag{217}$$

$$\dot{\tau}_{g_{max}} = \frac{2R_\oplus\omega}{c} = 3.1 \ \mu\text{sec/sec}. \tag{218}$$

If we observe at 22 GHz, equations (208) and (218) implies that the expected correlation may oscillate with frequency as high as 22 GHz $\times$ 3.1 $\times$ 10^{-6} = 68.2 kHz, or 68,200 cycles per second(!), due to the diurnal motion of the source across the fringe pattern. Indeed, no meaningful integration of the product of voltages is possible, unless we stop such an oscillation.

The ranges of variations of geometric delay and its rate, as examined above, and resulting variations of the phase in the cross-power spectrum of voltages, such as shown in equation (200), are orders of magnitude larger, than those caused by other delay effects (clock offset, atmospheric propagation delay, cable delay, etc.), and by other phase terms (LO initial phase, visibility phase, phase of bandpass characteristics, etc.). Therefore, the delay tracking and fringe stopping are, primarily, means for compensation of the geometric delay. Although some other, much smaller, effects are also taken into account in highly precise delay tracking and fringe stopping, performed in VLBI, we will postpone discussion of this problem, until we consider correlation processings in VLBI.

We use an appropriate theoretical prediction of the time–variable geometric delay, in order to determine length and rate of the instrumental delay, to be applied to actual delay tracking and fringe stopping. Such a theoretical prediction itself, for the geometric delay τ_{g_0}, has been traditionally denoted as τ_i, and called "instrumental delay". Please do not mix this up with unwanted "instrumentally induced delay" such as the one due to extension of transmission cables.

Let us estimate accuracies required to an instrumental delay τ_i and its rate $\dot{\tau}_i$.

In order to get a white fringe near the peak of the bandwidth pattern, a difference $\Delta\tau_g$ between the actual geometric delay τ_{g_0} and the predicted instrumental delay τ_i, which we called the "residual delay":

$$\Delta\tau_g = \tau_{g_0} - \tau_i, \tag{219}$$

must be kept well smaller than the coherence interval $\Delta\tau_B = 2/B$, where B is an observing bandwidth:

$$\Delta\tau_g \ll \Delta\tau_B = \frac{2}{B}. \tag{220}$$

Therefore, the acuuracy of the instrumental delay must be much better than 1 μsec when $B = 2$ MHz, and much better than 1 nsec when $B = 2$ GHz, as Table 1 shows. We now compare these required accuracies with maximum delay values $\tau_{g_{max}}$, estimated above for 300 m baseline connected–element radio interferometer (CERI), and for VLBI. Table 3 shows relative accuracy, in terms of $\Delta\tau_g/\tau_{g_{max}}$, required to the instrumental delay. The Table shows

Table 3: Relative accuracy of $\Delta\tau_g/\tau_{g_{max}}$ required to the instrumental delay.

	300 m CERI $\tau_{g_{max}} = 1\mu$sec	VLBI $\tau_{g_{max}} = 21$ msec
$B = 2$ MHz ($\Delta\tau_g \ll 1$ μsec)	$\ll 1$	$\ll 5 \times 10^{-5}$
$B = 2$ GHz ($\Delta\tau_g \ll 1$ nsec)	$\ll 10^{-3}$	$\ll 5 \times 10^{-8}$

that the required accuracy is rather high, for VLBI.

On the other hand, in order to carry out integration of the product of voltages in a correlator, for a duration of time sufficient to get high enough S/N ratio, we must apply an accurate theoretical model $\tau_i(t)$ to compensate for time variation of the geometric delay $\tau_{g_0}(t)$. If integration time is τ_a and observing frequency is ν, a requirement to $\Delta\dot{\tau}_g = (\dot{\tau}_{g_0} - \dot{\tau}_i)$ for successful integration, is conventionally given by

$$2\pi\nu\Delta\dot{\tau}_g\tau_a \leq 1. \tag{221}$$

This requirement corresponds to a condition that an accumulated change of

Table 4: Required accuracy of $\Delta\dot{\tau}_g$ and relative accuracy of $\Delta\dot{\tau}_g/\dot{\tau}_{g_{max}}$ (in parenthes) for a 300 m baseline CERI and a VLBI.

	300 m CERI $\tau_a = 1000$ s, $\dot{\tau}_{g_{max}} = 70$ ps/s	VLBI $\tau_a = 1$ s , $\dot{\tau}_{g_{max}} = 3.1\mu$s/s
$\nu = 10$ GHz	$\leq 1.6 \times 10^{-2}$ ps/s ($\leq 2.3 \times 10^{-4}$)	≤ 16 ps/s ($\leq 5.2 \times 10^{-6}$)
$\nu = 100$ GHz	$\leq 1.6 \times 10^{-3}$ ps/s ($\leq 2.3 \times 10^{-5}$)	≤ 1.6 ps/s ($\leq 5.2 \times 10^{-7}$)

the phase of the cross–power spectrum of voltages, such as shown in equation

(213), or of the phase in cosine terms of the expected correlation in equation (214), does not exceed 1 radian during the integration time τ_a.

Table 4 shows required accuracy of the residual delay rate $\Delta\dot{\tau}_g$, according to equation (221), and its ratio to the maximum delay rate, estimated above, for a 300 m baseline CERI with a typical integration time $\tau_a = 1000$ s, and a VLBI with a typical hardware integration time with a VLBI correlator $\tau_a = 1$ s, in two cases of observing frequency, $\nu = 10$ GHz and $\nu = 100$ GHz.

The requred accuracies are fairly high, especially in VLBI. Therefore, theoretical prediction of the instrumental delay is usually based on state–of–the–art geophysical and astronomical models of station coordinates, radio source coordinates, and irregularities in the rotational motion of the Earth.

2.8.4 Insertion of Instrumental Delay at IF Band

Now let us consider the case, when a time–variable instrumental delay τ_i is inserted in the IF–band signal transmission system of antenna 2, as illustrated in Figure 53.

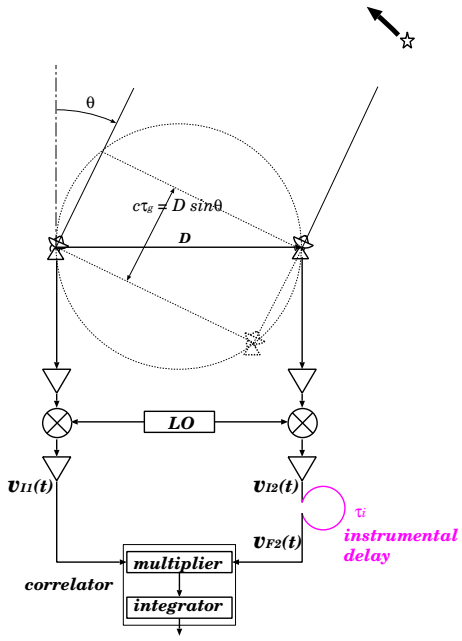

Figure 53: Time–variable instrumental delay inserted at IF band.

After the insertion of the delay circuit, IF voltage signal $v_{F2}(t)$, and its Fourier transform $\tilde{v}_{F2}(\omega)$, are given by

$$v_{F2}(t) = v_{I2}(t - \tau_i), \quad \text{and, therefore,} \quad \tilde{v}_{F2}(\omega) = \tilde{v}_{I2}(\omega)\, e^{-i\omega\tau_i}, \qquad (222)$$

where $v_{I2}(t)$, and $\tilde{v}_{I2}(\omega)$, correspond to the IF voltage signal before the instrumental delay. We again used the shift theorem given in equation (69), for deriving equation (222).

Following discussions in subsection 2.8.1, we take a cross–correlation of Fourier transforms $\tilde{v}_{I1}(\omega)$ and $\tilde{v}_{F2}(\omega)$ of IF voltage signals, to obtain

$$\langle \tilde{v}_{I1}(\omega)\, \tilde{v}^*_{F2}(\omega')\rangle = \langle \tilde{v}_{I1}(\omega)\, \tilde{v}^*_{I2}(\omega')\rangle\, e^{i\omega'\tau_i}. \tag{223}$$

Since the IF signals $v_{I1}(t)$ and $v_{F2}(t)$, as well as $v_{I1}(t)$ and $v_{I2}(t)$, are assumed to be jointly stationary random processes, the cross–correlations of their Fourier transforms and cross–power spectra are related to each other, by equation (84), as:

$$\begin{aligned} \langle \tilde{v}_{I1}(\omega)\, \tilde{v}^*_{F2}(\omega')\rangle &= 2\pi S_{v_{I1}v_{F2}}(\omega)\, \delta(\omega-\omega'), \\ \langle \tilde{v}_{I1}(\omega)\, \tilde{v}^*_{I2}(\omega')\rangle &= 2\pi S_{v_{I1}v_{I2}}(\omega)\, \delta(\omega-\omega'). \end{aligned} \tag{224}$$

Comparing equations (223) and (224), we see that the cross–power spectrum $S_{v_{I1}v_{F2}}(\omega)$ of the IF voltage signal $v_{I1}(t)$ of antenna 1 and the delay–inserted IF voltage signal $v_{F2}(t)$ of antenna 2, and the cross–power spectrum $S_{v_{I1}v_{I2}}(\omega)$ of the IF voltages without delay insertion, are related to each other as:

$$S_{v_{I1}v_{F2}}(\omega) = S_{v_{I1}v_{I2}}(\omega)\, e^{i\omega\tau_i}. \tag{225}$$

Inserting equation (197) to this equation, we have

$$\begin{aligned} &S_{v_{I1}v_{F2}}(\omega) \\ &= \frac{1}{4}\, A_0\, \{e^{-i[\omega_{LO}\,\tau_{g_0}+\omega\Delta\tau_g+\phi_{LO1}-\phi_{LO2}]}\, \mathcal{V}(\omega_{LO}+\omega)\, H_1(\omega)\, H_2^*(\omega) \\ &\quad + e^{i[\omega_{LO}\,\tau_{g_0}-\omega\Delta\tau_g+\phi_{LO1}-\phi_{LO2}]}\, \mathcal{V}^*(\omega_{LO}-\omega)\, H_1^*(-\omega)\, H_2(-\omega)\}, \end{aligned} \tag{226}$$

where, $H_1(\omega)$ and $H_2(\omega)$ are the system functions for responses of combined RF and IF filters, ω_{LO} is the local oscillator frequency, $\mathcal{V}(\omega_{LO}\pm\omega)$ is the complex visibility of the source, A_0 is the geometric mean of effective apertures of element antennas, and again, $\Delta\tau_g = \tau_{g_0} - \tau_i$ is the residual delay.

Contrary to equation (213), which describes the effect of the delay insertion at the RF band, we cannot here completely compensate for the geometric delay τ_{g_0}, even if we insert an instrumental delay which is exactly equal to the geometric delay $\tau_i = \tau_{g_0}$. In fact, the $\omega\tau_{g_0}$ term, where ω is the IF frequency, is compensated by the $\omega\tau_i$ and is reduced to $\omega\Delta\tau_g$, but the $\omega_{LO}\tau_{g_0}$ term is still left uncompensated in the exponential functions of equation (226).

In order to see this point more clearly, let us derive the expected correlation $R_{v_{I1}v_{F2}}(0)$, as we did in equation (207), assuming again a rectangular

filter, given in equation (205), and constant visibility over receiving bandwidth, given in equation (206). Since everything is the same as given in equation (207), except for a coefficient of ω term in the exponential function of the integrand, which is now $\omega\Delta\tau_g$ instead of $\omega\tau_{g_0}$, we obtain:

$$\begin{aligned} R_{v_{I1}v_{F2}}(0) = & \frac{A_0 GB \mid \mathcal{V} \mid}{2} \frac{\sin(\pi B\Delta\tau_g)}{\pi B\Delta\tau_g} \\ & \times \quad [\cos(\omega_{LO}\tau_{g_0} + \omega_I\Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v) \\ & + \quad \cos(\omega_{LO}\tau_{g_0} - \omega_I\Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v)], \end{aligned} \tag{227}$$

where ω_I is the IF band center frequency as before, and other notations are also the same as those in equation (208). This equation is quite similar to equation (208) if we replace τ_{g_0} by $\Delta\tau_g$, except however for the $\omega_{LO}\tau_{g_0}$ terms in cosine functions.

It is evident, from this equation, that insertion of an instrumental delay τ_i at the IF band, and resultant compensation of the $\omega\tau_{g_0}$ term, enables us to detect a white fringe at the center of the coherence interval. However, the $\omega_{LO}\tau_{g_0}$ terms in arguments of cosine functions are left uncompensated, which will cause the enormously rapid oscillation of the expected correlation. Therefore, insertion of an instrumental delay at IF band can perform the delay tracking, **but not** the fringe stopping.

2.8.5 Separation of Delay Tracking and Fringe Stopping Due to Frequency Conversion

The reason, why the insertion of the instrumental delay at IF band can perform only the delay tracking, but not the fringe stopping, may be easily understood by examining Figure 54. This figure illustrates a time variation of the phase spectrum of IF voltages shown in Figure 46, in a particular case of a single–sideband reception of USB component.

Let us consider phase spectra in cross–power spectra of voltage signals without delay insertion, and compare the phase spectrum in IF band with the one in RF band. In doing so, we ignore minor phase terms, such as LO initial phase, visibility phase, and phase of bandpass characteristics, which figure in equation (200), leaving only the main term due to the geometric delay.

Then, the phase spectrum in RF band is a straight line crossing the origin, with inclination equal to the geometric delay τ_{g_0}, as shown in equation (103) and Figure 21 for a simple interferometer model, or in equations (168), and (177), for a more realistic interferometer.

However, after the frequency conversion, the phase spectrum in the IF band no longer crosses the origin, since the frequency conversion shifts the band–limited spectrum at RF band to IF band without changing its shape, as shown in Figures 44, 46, and 54.

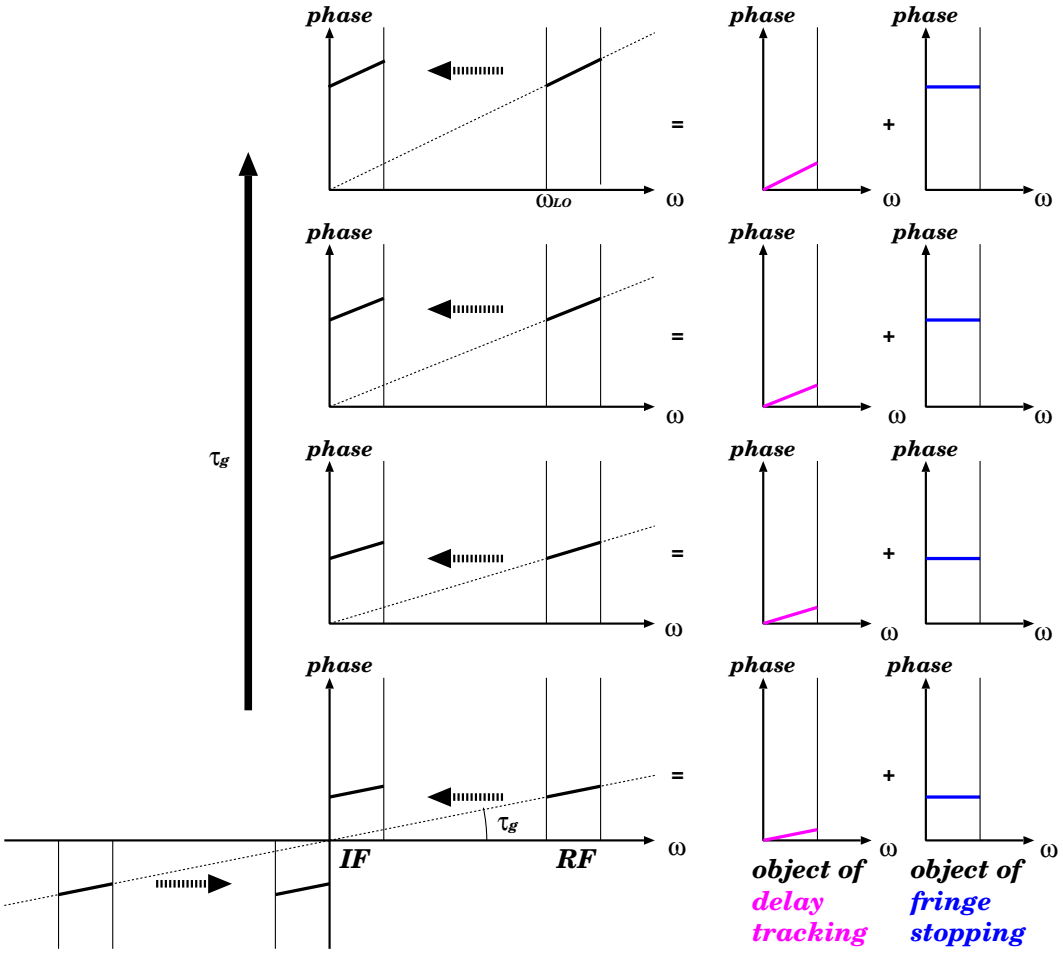

Figure 54: Necessity of fringe stopping, besides delay tracking in IF band. We assume here a USB single–sideband reception, and show only positive frequency range of the phase spectrum, except in the bottom panel.

Therefore, the frequency converted phase spectrum in IF band is now represented by a sum of two components, namely an inclined straight line $\omega\tau_{g0}$, crossing the origin, **plus** a frequency–independent phase shift equal to $\omega_{LO}\tau_{g0}$, where ω_{LO} is the LO frequency, as illustrated in the right panels of Figure 54 and in equation (200). This phase shift rapidly changes in time, with the speed as high as several tens kHz (as we saw in an example of an

intercontinental VLBI observation at 22 GHz in subsection 2.8.3), due to the diurnal variation of the geometric delay τ_{g_0}.

The insertion of the instrumental delay at the IF band compensates for the first component, i.e. reduces the phase slope across the frequency band to nearly zero, and thus makes it possible to get a white fringe in the middle of the coherence interval, as evident from equation (208). However, the instrumental delay in the IF band does not compensate for the second component at all. The second component, if left uncompensated, would cause the enormous oscillation of the expected correlation and make it impossible to integrate the product of voltages for achieving high enough signal to noise ratio. Therefore, we still need to compensate for the rapid phase shift in the second component, by somehow controlling the interferometer phase. This operation is the "fringe stopping", which is separated from the "delay tracking" after the frequency conversion.

2.8.6 Actual Implementations of Delay Tracking

In early history of radio interferometry, analog delay cables, as schematically

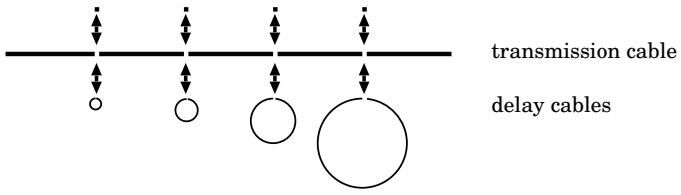

Figure 55: Analog delay cables used for delay tracking in early history of radio interferometry.

illustrated in Figure 55, were used for delay tracking. N delay cables with lengths τ_0, $2\tau_0$, $4\tau_0$, $8\tau_0$, $\cdots$, $2^{N-1}\tau_0$, were mechanically inserted into, and removed from, a signal transmission cable of an interferometer, to realize a time variable instrumental delay τ_i with a range of variation: $\tau_0 \leq \tau_i \leq (2^N - 1)\tau_0$. Of cource, the longest delay cable must have a length as large as about a half of baseline length of the interferometer. Therefore, such analog delay cables could be used for connected–element interferometers with baseline lengths of several hundreds meters, but definitely not for VLBIs.

Recently, most of radio interferometers, including VLBIs, use digital delay circuits for delay tracking. In order to use this technique, data signal is first digitized, and then fed to a bulk memory, which is a kind of "ring buffer", schematically shown in Figure 56.

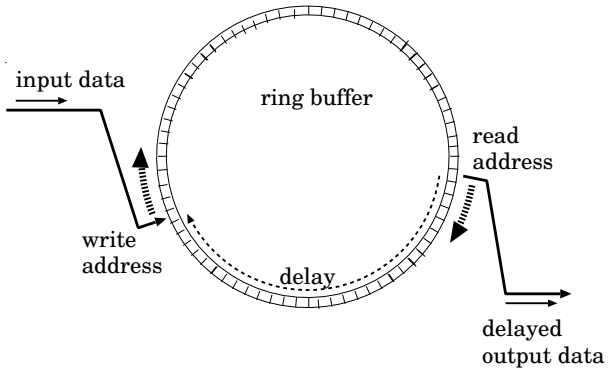

Figure 56: A "ring buffer" used for digital delay tracking.

Data are written to, and read from, the ring buffer at different addresses, and these write and read addresses are incremented by one at each "clock timing pulse" of digital circuit. An interval of succcessive clock timing pulses corresponds to a sampling interval t_s of digitization of the data. Thus, if the difference between the write and read addresses in the ring buffer is n, an instrumental delay equal to nt_s is realized in this way. And further shifting the write (or read) address periodically, once per a certain number of clock timing pulses, we can vary this instrumental delay in time. We will examine this digital delay tracking in more detail, when we will discuss VLBI correlators.

2.8.7 Actual Implementations of Fringe Stopping

In connected–element interferometers, the fringe stopping is often carried out by actively controlling the initial phases ϕ_{LO1} and ϕ_{LO2} of the LO reference signals, as illustrated in Figure 57, so that

$$\phi_{LO1}(t) - \phi_{LO2}(t) = -\omega_{LO}\,\tau_i(t) + \psi_{LO1} - \psi_{LO2}, \tag{228}$$

where $\psi_{LO1} - \psi_{LO2}$ is a nearly constant part of the difference of the initial phases of the LO signals, and τ_i is the instrumental delay, i.e., the theoretical prediction of the geometric delay τ_{g_0}.

Assuming quasi–static LO initial phases, similarly to what we did in subsection 2.4 for the geometric delay, we can just insert equation (228) to equation (226) for the cross–power spectrum of IF voltages, and to equation (227) for the expected correlation in the case of a continuum flat spectrum

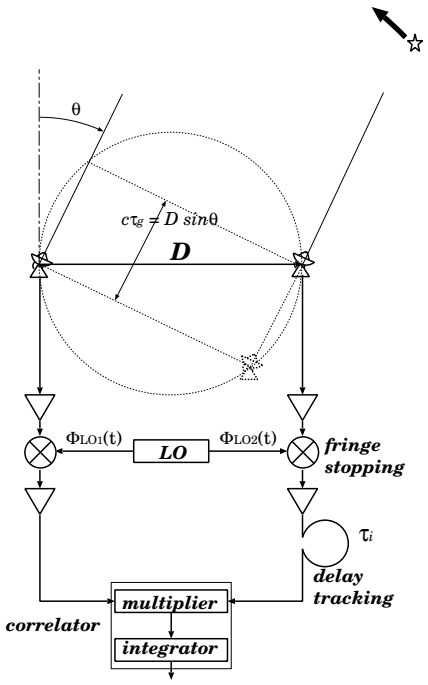

Figure 57: Fringe stopping by active control of LO initial phases.

source and rectangular filters, after the delay tracking at IF band. Then, we obtain the cross–power spectrum of IF voltages:

$$
\begin{aligned}
& S^f_{v_{I1}v_{F2}}(\omega) \\
&= \frac{1}{4}\, A_0\, \{ e^{-i[\omega_{LO}\,\Delta\tau_g + \omega\Delta\tau_g + \psi_{LO1} - \psi_{LO2}]}\, \mathcal{V}(\omega_{LO}+\omega)\, H_1(\omega)\, H_2^*(\omega) \\
&\qquad + e^{i[\omega_{LO}\,\Delta\tau_g - \omega\Delta\tau_g + \psi_{LO1} - \psi_{LO2}]}\, \mathcal{V}^*(\omega_{LO}-\omega)\, H_1^*(-\omega)\, H_2(-\omega)\},
\end{aligned}
\tag{229}
$$

and expected correlation:

$$
\begin{aligned}
R^f_{v_{I1}v_{F2}}(0) = \frac{A_0 GB \mid \mathcal{V} \mid}{2} & \frac{\sin(\pi B \Delta\tau_g)}{\pi B \Delta\tau_g} \\
\times \quad & [\cos(\omega_{LO}\Delta\tau_g + \omega_I \Delta\tau_g + \psi_{LO1} - \psi_{LO2} - \Phi_v) \\
+ \quad & \cos(\omega_{LO}\Delta\tau_g - \omega_I \Delta\tau_g + \psi_{LO1} - \psi_{LO2} - \Phi_v)],
\end{aligned}
\tag{230}
$$

respectively.

Now we can stop the rapid phase changes in the cross–power spectrum, as well as in the expected correlation, provided that we can always keep the

residual delay to be nearly zero:

$$\Delta\tau_g = \tau_{g\{0\}} - \tau_i \cong 0.$$

Note that this active control of LO initial phases can stop the phase changes both in USB and LSB components of the cross–power spectrum and expected correlation of IF voltages, as evidenced by equations (229) and (230). In particular, arguments of first and second cosine terms in the RHS of equation (230), which show USB and LSB contributions, respectively, are both stopped by this operation.

Another method for fringe stopping, which has been successfully applied in particular to VLBI digital correlators uses multiplication of a sinusoidal function of time to an IF voltage, as shown in Figure 58.

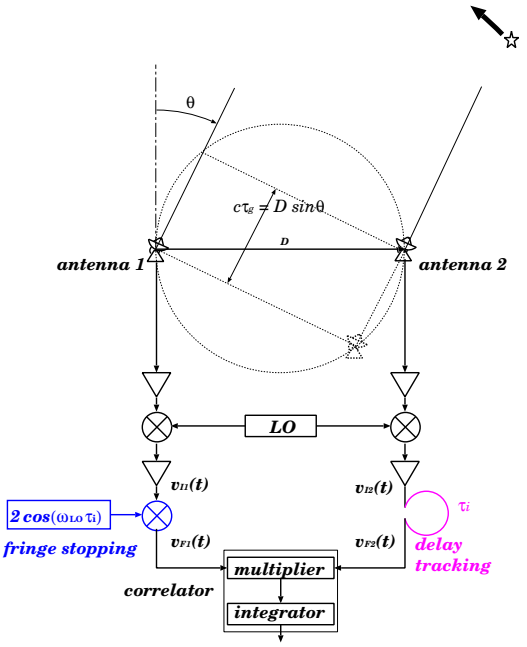

Figure 58: Fringe stopping by multiplication of a sinusoidal function.

In this method, a sinusoidal function

$$2\cos(\omega_{LO}\tau_i), \tag{231}$$

where ω_{LO} is the LO frequency, and τ_i is the theoretically predicted instrumental delay, is multiplied to IF voltage $v_{I1}(t)$ of antenna 1.

Assuming again that $\omega_{LO}\tau_i$ is a quasi–static quantity, we obtain for a new IF voltage $v_{F1}(t)$ of antenna 1 after the multiplication of $2\cos(\omega_{LO}\tau_i)$:

$$v_{F1}(t) = v_{I1}(t)2\cos(\omega_{LO}\tau_i) = v_{I1}(t)(e^{i\omega_{LO}\tau_i} + e^{-i\omega_{LO}\tau_i}), \tag{232}$$

and for its Fourier transform $\tilde{v}_{F1}(\omega)$:

$$\tilde{v}_{F1}(\omega) = \tilde{v}_{I1}(\omega)(e^{i\omega_{LO}\tau_i} + e^{-i\omega_{LO}\tau_i}), \tag{233}$$

where $\tilde{v}_{I1}(\omega)$ is a Fourier transform of $v_{I1}(t)$. Then we have a cross–correlation of Fourier transforms of IF voltages:

$$\langle\tilde{v}_{F1}(\omega)\tilde{v}^*_{F2}(\omega')\rangle = \langle\tilde{v}_{I1}(\omega)\tilde{v}^*_{F2}(\omega')\rangle(e^{i\omega_{LO}\tau_i} + e^{-i\omega_{LO}\tau_i}), \tag{234}$$

where $\tilde{v}_{F2}(\omega)$ is the IF voltage of antenna 2 after the delay insertion. Assuming jointly stationary random processes, we obtain for a cross–power spectrum of IF voltages $v_{F1}(t)$ and $v_{F2}(t)$:

$$S_{v_{F1}v_{F2}}(\omega) = S_{v_{I1}v_{F2}}(\omega)(e^{i\omega_{LO}\tau_i} + e^{-i\omega_{LO}\tau_i}), \tag{235}$$

where $S_{v_{I1}v_{F2}}(\omega)$ is the cross–power spectrum of IF voltages $v_{I1}(t)$ and $v_{F2}(t)$, which is given in equation (226). Then, inserting equation (226) to equation (235), we have

$$\begin{aligned} S_{v_{F1}v_{F2}}(\omega) = \frac{1}{4}A_0 \\ \times\{&e^{-i[\omega_{LO}\,\Delta\tau_g+\omega\Delta\tau_g+\phi_{LO1}-\phi_{LO2}]}\,\mathcal{V}(\omega_{LO}+\omega)\,H_1(\omega)\,H_2^*(\omega) \\ &+e^{-i[\omega_{LO}(\tau_{g_0}+\tau_i)+\omega\Delta\tau_g+\phi_{LO1}-\phi_{LO2}]}\,\mathcal{V}(\omega_{LO}+\omega)\,H_1(\omega)\,H_2^*(\omega) \\ &+e^{i[\omega_{LO}\,\Delta\tau_g-\omega\Delta\tau_g+\phi_{LO1}-\phi_{LO2}]}\,\mathcal{V}^*(\omega_{LO}-\omega)\,H_1^*(-\omega)\,H_2(-\omega) \\ &+e^{i[\omega_{LO}(\tau_{g_0}+\tau_i)-\omega\Delta\tau_g+\phi_{LO1}-\phi_{LO2}]}\,\mathcal{V}^*(\omega_{LO}-\omega)\,H_1^*(-\omega)\,H_2(-\omega)\}, \end{aligned} \tag{236}$$

where $\Delta\tau_g = \tau_{g_0} - \tau_i$. In first and third terms in the RHS of this equation, rapid oscillations of exponential terms are almost stopped, as long as $\tau_i \cong \tau_{g_0}$. But, in second and fourth terms in the RHS, even more rapid oscillations with almost doubled frequency $\omega_{LO}(\dot{\tau_{g_0}} + \dot{\tau}_i)$ remain. Contributions of these rapidly oscillating terms almost disappear when the product of IF voltages are integrated for some duration of time in a correlator. Therefore, we have an "effective" cross–power spectrum of IF voltages,

$$\begin{aligned} S_{v_{F1}v_{F2}}(\omega) = \frac{1}{4}A_0 \\ \times\{&e^{-i[\omega_{LO}\,\Delta\tau_g+\omega\Delta\tau_g+\phi_{LO1}-\phi_{LO2}]}\,\mathcal{V}(\omega_{LO}+\omega)\,H_1(\omega)\,H_2^*(\omega) \\ &+e^{i[\omega_{LO}\,\Delta\tau_g-\omega\Delta\tau_g+\phi_{LO1}-\phi_{LO2}]}\,\mathcal{V}^*(\omega_{LO}-\omega)\,H_1^*(-\omega)\,H_2(-\omega)\}, \end{aligned} \tag{237}$$

which has essentially the same form as the one given in equation (229).

Note that, in this simple method of multiplication of a sinusoidal function to an IF voltage, phase drifts are stopped again in both USB and LSB components.

2.9 Correlator Outputs

Now we proceed to the final element of processing units in our realistic interferometer, the correlator (Figure 59).

2.9.1 Multiplier and Integrator

As we saw earlier, a key component of a correlator in an interferometer is a device with two inputs and one output, which is composed of a multiplier and an integrator (right panel of Figure 59).

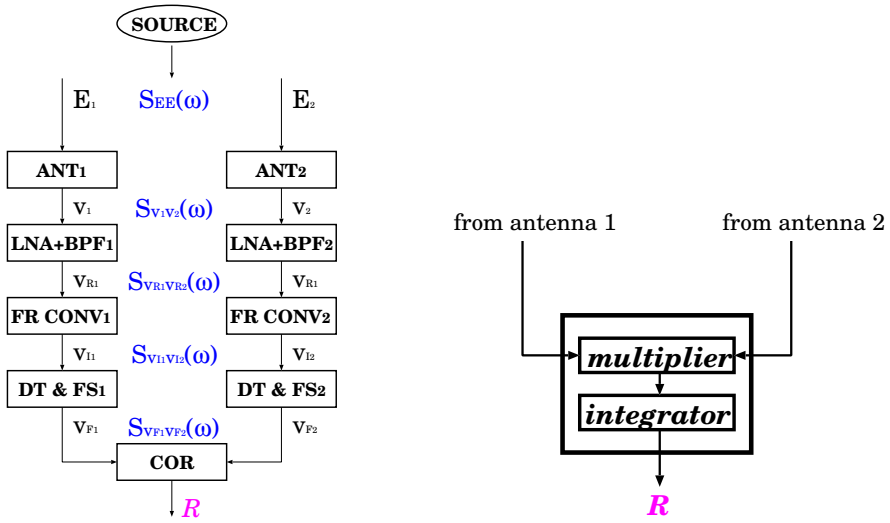

Figure 59: The output of a correlator.

The expected correlation, which is the statistical expectation of the product of the voltages, must be contained as a signal component in a multiplier output of such a correlator. Since the expected correlation does not rapidly oscillate any more after the delay tracking and fringe stopping, we can integrate (time–average) the multiplier output during some time to suppress the noise component, and detect this nearly constant expected correlation itself with a high signal to noise ratio. Therefore, an output of an actual correlator

can be well approximated by the expected correlation after the delay tracking and fringe stopping. The expected correlation is easily derived from the cross–power spectrum of the IF voltages by an inverse Fourier transformation with argument $\tau = 0$, as we showed in equation (204).

Now we have slightly different notations for the cross–power spectrum of IF voltages in equation (229) and in equation (237), depending on practical implementations of the fringe stopping, though they are essentially equivalent to each other. We adopt, hereafter, the expression given in equation (237), which is familiar in VLBI practice, for describing the cross–power spectrum after the delay tracking and fringe stopping.

Then, the expected correlator output, or the expected correlation, is derived from equation (237) by the inverse Fourier transformation with $\tau = 0$:

$$R_{v_{F1}v_{F2}}(0) = \frac{A_0}{4\pi}\,\Re\left[e^{-i(\omega_{LO}\Delta\tau_g+\phi_{LO1}-\phi_{LO2})}\int_0^\infty \mathcal{V}(\omega_{LO}+\omega)\,e^{-i\omega\Delta\tau_g}\,H_1(\omega)\,H_2^*(\omega)\,d\omega\right.$$
$$\left. + e^{i(\omega_{LO}\Delta\tau_g+\phi_{LO1}-\phi_{LO2})}\int_0^\infty \mathcal{V}^*(\omega_{LO}-\omega)\,e^{-i\omega\Delta\tau_g}\,H_1^*(-\omega)\,H_2(-\omega)\,d\omega\right]. \tag{238}$$

In the special case of a continuum flat spectrum source and rectangular filters, the expected correlation will have essentially the same form as the one given in equation (230):

$$\begin{aligned} R_{v_{F1}v_{F2}}(0) = {} & \frac{A_0 GB \mid \mathcal{V} \mid}{2}\,\frac{\sin(\pi B\Delta\tau_g)}{\pi B\Delta\tau_g} \\ & \times \quad [\cos(\omega_{LO}\Delta\tau_g + \omega_I\Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v) \\ & + \quad \cos(\omega_{LO}\Delta\tau_g - \omega_I\Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v)]. \end{aligned} \tag{239}$$

We will use these forms of the expected correlation given in equations (238) and (239) as theoretical expressions of the correlator outputs.

In VLBI, no operation for delay tracking and fringe stopping is done, usually, in observing stations. Instead, they are done in VLBI correlators. Therefore, a VLBI correlator is a little more complicated than the mere multiplier and integrator, as we will see later.

2.9.2 Single Sideband (USB or LSB) Reception

So far, we considered a case of double sideband (DSB) reception, i.e. a case when we receive signals contained in both upper sideband (USB) and lower sideband (LSB) of LO frequency in the original RF band spectrum. In this

case, the USB and LSB contributions are superposed in IF band spectrum, as we saw earlier.

The DSB reception is not convenient for observing spectralline sources, since different lines in USB and LSB are mixed up in the same IF spectrum. Moreover, the DSB reception causes additional complications in data calibrations for continuum sources. Therefore, in VLBI, we mostly use a single sideband (SSB), i.e., either of USB or LSB, reception.

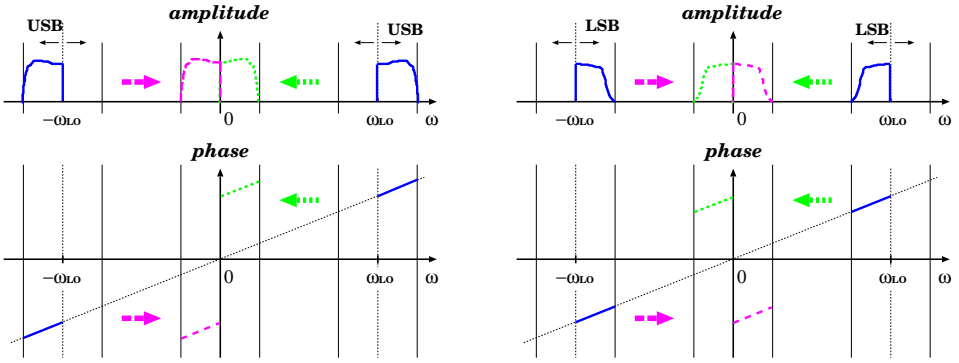

Figure 60: Amplitude and phase spectra of IF voltages in USB (left) and LSB (right) receptions. For simplicity, contributions of the geometric delay only are shown in the phase spectra.

Amplitude and phase spectra of IF voltages, which were shown in Figure 46 in the DSB reception case, now take simpler forms, in the SSB reception case, as shown in Figure 60.

Since, cross–power spectra of IF voltages both in USB and LSB receptions are Hermitian symmetric, as illustrated in Figure 60, we do not lose generality if we confine ourselves to consider only positive frequency range ($\omega \geq 0$) of the spectra.

From equation (197), we see that, in positive frequency range $\omega \geq 0$, cross–power spectra of IF voltages before the delay tracking and fringe stopping in USB and LSB receptions, respectively, are given by

$$S^{USB}_{v_{I1}v_{I2}}(\omega) = \frac{1}{4} A_0 \, e^{-i[\omega_{LO}\,\tau_{g0} + \omega\tau_{g0} + \phi_{LO1} - \phi_{LO2}]} \, \mathcal{V}(\omega_{LO} + \omega) \, H_1(\omega) \, H_2^*(\omega), \tag{240}$$

$$S^{LSB}_{v_{I1}v_{I2}}(\omega) = \frac{1}{4} A_0 \, e^{i[\omega_{LO}\,\tau_{g0} - \omega\tau_{g0} + \phi_{LO1} - \phi_{LO2}]} \, \mathcal{V}^*(\omega_{LO} - \omega) \, H_1^*(-\omega) \, H_2(-\omega), \tag{241}$$

which correspond to positive frequency sides ($\omega \geq 0$) of the USB and LSB spectra of Figure 60, as separately illustrated in Figure 61.

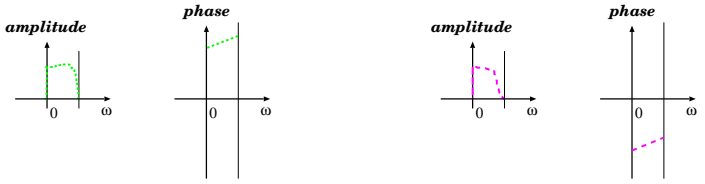

Figure 61: Amplitude and phase spectra of IF voltages in USB (left) and LSB (right) receptions in positive frequency range $\omega \geq 0$.

After the delay tracking and fringe stopping, they are reduced to

$$S^{USB}_{v_{F1}v_{F2}}(\omega) = \frac{1}{4}\, A_0\, e^{-i[\omega_{LO}\,\Delta\tau_g + \omega\Delta\tau_g + \phi_{LO1} - \phi_{LO2}]}\, \mathcal{V}(\omega_{LO}+\omega)\, H_1(\omega)\, H_2^*(\omega), \tag{242}$$

$$S^{LSB}_{v_{F1}v_{F2}}(\omega) = \frac{1}{4}\, A_0\, e^{i[\omega_{LO}\,\Delta\tau_g - \omega\Delta\tau_g + \phi_{LO1} - \phi_{LO2}]}\, \mathcal{V}^*(\omega_{LO}-\omega)\, H_1^*(-\omega)\, H_2(-\omega), \tag{243}$$

in the positive frequency range $\omega \geq 0$.

We can rewrite equations (240) and (241), and equations (242) and (243) as well, using amplitudes and phases of the complex visibility and bandpass characteristics, which we introduced in equations (198) and (199).

Before the delay tracking and fringe stopping, equation (240) for the USB spectrum of IF voltages in $\omega \geq 0$ becomes

$$S^{USB}_{v_{I1}v_{I2}}(\omega) = A_U(\omega)\, e^{-i\Phi_U(\omega,\,t)},$$

with amplitude:

$$A_U(\omega) = \frac{1}{4}\, A_0\, \mid \mathcal{V}(\omega_{LO}+\omega) \mid\mid H_1(\omega)\, H_2^*(\omega) \mid,$$

and phase:

$$\Phi_U(\omega,\, t) = (\omega_{LO}+\omega)\tau_{g_0} + \phi_{LO1} - \phi_{LO2} - \Phi_v(\omega_{LO}+\omega) - \Phi_b(\omega), \tag{244}$$

while equation (241) for the LSB spectrum of IF voltages in $\omega \geq 0$ becomes

$$S^{LSB}_{v_{I1}v_{I2}}(\omega) = A_L(\omega)\, e^{-i\Phi_L(\omega,\,t)},$$

with amplitude:

$$A_L(\omega) = \frac{1}{4}\, A_0\, \mid \mathcal{V}(\omega_{LO}-\omega) \mid\mid H_1(-\omega)\, H_2^*(-\omega) \mid,$$

and phase:

$$\Phi_L(\omega,\, t) = -(\omega_{LO}-\omega)\tau_{g_0} - \phi_{LO1} + \phi_{LO2} + \Phi_v(\omega_{LO}-\omega) + \Phi_b(-\omega). \tag{245}$$

After the delay tracking and fringe stopping, the USB spectrum in $\omega \geq 0$ becomes

$$S^{USB}_{v_{F1}v_{F2}}(\omega) = A_U(\omega)\, e^{-i\Delta\Phi_U(\omega,t)},$$

with amplitude:

$$A_U(\omega) = \frac{1}{4}\, A_0 \;\mid \mathcal{V}(\omega_{LO}+\omega) \mid\mid H_1(\omega)\, H_2^*(\omega) \mid,$$

and residual phase:

$$\Delta\Phi_U(\omega,\, t) = (\omega_{LO}+\omega)\Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v(\omega_{LO}+\omega) - \Phi_b(\omega), \tag{246}$$

while the LSB spectrum in $\omega \geq 0$ becomes

$$S^{LSB}_{v_{F1}v_{F2}}(\omega) = A_L(\omega)\, e^{-i\Delta\Phi_L(\omega,t)},$$

with amplitude:

$$A_L(\omega) = \frac{1}{4}\, A_0 \;\mid \mathcal{V}(\omega_{LO}-\omega) \mid\mid H_1(-\omega)\, H_2^*(-\omega) \mid,$$

and residual phase:

$$\Delta\Phi_L(\omega,\, t) = -(\omega_{LO}-\omega)\Delta\tau_g - \phi_{LO1} + \phi_{LO2} + \Phi_v(\omega_{LO}-\omega) + \Phi_b(-\omega), \tag{247}$$

where $\Delta\tau_g$ is the residual delay: $\Delta\tau_g = \tau_{g_0} - \tau_i$.

2.9.3 Correlator Outputs in Single Sideband Reception

As we discussed earlier, correlator outputs are closely approximated by the expected correlations after the delay tracking and fringe stopping. Therfore, in view of equation (238) or equations (242) and (243), theoretical expressions for the correlator outputs in USB and LSB receptions are given by

$$R^{USB}_{v_{F1}v_{F2}}(0) = \frac{A_0}{4\pi}\, \Re[e^{-i(\omega_{LO}\Delta\tau_g + \phi_{LO1} - \phi_{LO2})} \int_0^\infty \mathcal{V}(\omega_{LO}+\omega)\, e^{-i\omega\Delta\tau_g}\, H_1(\omega)\, H_2^*(\omega)\, d\omega], \tag{248}$$

and

$$R^{LSB}_{v_{F1}v_{F2}}(0) = \frac{A_0}{4\pi}\, \Re[e^{i(\omega_{LO}\Delta\tau_g + \phi_{LO1} - \phi_{LO2})} \int_0^\infty \mathcal{V}^*(\omega_{LO}-\omega)\, e^{-i\omega\Delta\tau_g}\, H_1^*(-\omega)\, H_2(-\omega)\, d\omega], \tag{249}$$

respectively.

If we assume the case of rectangular filters and constant visibility, the above equations are reduced to

$$R^{USB}_{v_{F1}v_{F2}}(0) = \frac{A_0 GB \mid \mathcal{V} \mid}{2} \frac{\sin(\pi B \Delta\tau_g)}{\pi B \Delta\tau_g} \cos(\omega_{LO}\Delta\tau_g + \omega_I \Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v), \tag{250}$$

and

$$R^{LSB}_{v_{F1}v_{F2}}(0) = \frac{A_0 GB \mid \mathcal{V} \mid}{2} \frac{\sin(\pi B \Delta\tau_g)}{\pi B \Delta\tau_g} \cos(\omega_{LO}\Delta\tau_g - \omega_I \Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v). \tag{251}$$

For a more general case of an arbitrary filter shape, when the complex visibitlity is assumed to be constant only in each of USB and LSB receiving bandwidths, and the filters, with a passband:

$$\omega_I - \Delta\omega/2 \leq \mid \omega \mid \leq \omega_I + \Delta\omega/2,$$

are not necessarily rectangular, we have in the USB reception

$$\int_0^\infty \mathcal{V}(\omega_{LO}+\omega)\, e^{-i\omega\Delta\tau_g}\, H_1(\omega)\, H_2^*(\omega)\, d\omega = \mathcal{V}(\omega_{LO}+\omega) \int_{\omega_I - \frac{\Delta\omega}{2}}^{\omega_I + \frac{\Delta\omega}{2}} e^{-i\omega\Delta\tau_g}\, H_1(\omega)\, H_2^*(\omega)\, d\omega$$

$$= \mathcal{V}(\omega_{LO}+\omega)\, e^{-i\omega_I\Delta\tau_g} \int_{-\frac{\Delta\omega}{2}}^{\frac{\Delta\omega}{2}} e^{-i\omega'\Delta\tau_g}\, H_1(\omega_I+\omega')\, H_2^*(\omega_I+\omega')\, d\omega', \tag{252}$$

and in the LSB reception

$$\int_0^\infty \mathcal{V}^*(\omega_{LO}-\omega)\, e^{-i\omega\Delta\tau_g}\, H_1^*(-\omega)\, H_2(-\omega)\, d\omega = \mathcal{V}^*(\omega_{LO}-\omega) \int_{\omega_I - \frac{\Delta\omega}{2}}^{\omega_I + \frac{\Delta\omega}{2}} e^{-i\omega\Delta\tau_g}\, H_1^*(-\omega)\, H_2(-\omega)\, d\omega$$

$$= \mathcal{V}^*(\omega_{LO}-\omega)\, e^{-i\omega_I\Delta\tau_g} \int_{-\frac{\Delta\omega}{2}}^{\frac{\Delta\omega}{2}} e^{-i\omega'\Delta\tau_g}\, H_1^*(-\omega_I-\omega')\, H_2(-\omega_I-\omega')\, d\omega'. \tag{253}$$

Now, let us introduce "bandwidth patterns" $\mathcal{B}_{12}$ defined by

$$\mathcal{B}^{USB}_{12}(B, \Delta\tau_g) = \frac{1}{4\pi} \int_{-\frac{\Delta\omega}{2}}^{\frac{\Delta\omega}{2}} e^{-i\omega'\Delta\tau_g}\, H_1(\omega_I+\omega')\, H_2^*(\omega_I+\omega')\, d\omega', \tag{254}$$

$$\mathcal{B}^{LSB}_{12}(B, \Delta\tau_g) = \frac{1}{4\pi} \int_{-\frac{\Delta\omega}{2}}^{\frac{\Delta\omega}{2}} e^{-i\omega'\Delta\tau_g}\, H_1^*(-\omega_I-\omega')\, H_2(-\omega_I-\omega')\, d\omega', \tag{255}$$

for USB and LSB receptions, respectively. The suffix '12' here accentuates that this pattern comes from bandpass characteristics of receiving systems in antennas 1 and 2. Then, if we represent complex visibilities and bandwidth patterns through their amplitudes and phases:

$$\mathcal{V}(\omega_{LO}+\omega_I) =\mid \mathcal{V}^U \mid e^{i\Phi^U{}_v} \quad \text{and} \quad \mathcal{V}(\omega_{LO}-\omega_I) =\mid \mathcal{V}^L \mid e^{i\Phi^L{}_v}, \tag{256}$$

and

$$\mathcal{B}_{12}^{USB}(B,\Delta\tau_g) =\mid \mathcal{B}_{12}^U \mid e^{i\Phi^U{}_B} \quad \text{and} \quad \mathcal{B}_{12}^{LSB}(B,\Delta\tau_g) =\mid \mathcal{B}_{12}^L \mid e^{i\Phi^L{}_B}, \tag{257}$$

equations (248) and (249) for theoretical expressions for correlator outputs are reduced to

$$R^{USB}_{v_{F1}v_{F2}}(0) = A_0 \mid \mathcal{V}^U \mid\mid \mathcal{B}_{12}^U \mid \cos(\omega_{LO}\Delta\tau_g+\omega_I\Delta\tau_g+\phi_{LO1}-\phi_{LO2}-\Phi_v^U-\Phi_B^U), \tag{258}$$

and

$$R^{LSB}_{v_{F1}v_{F2}}(0) = A_0 \mid \mathcal{V}^L \mid\mid \mathcal{B}_{12}^L \mid \cos(\omega_{LO}\Delta\tau_g-\omega_I\Delta\tau_g+\phi_{LO1}-\phi_{LO2}-\Phi_v^L+\Phi_B^L), \tag{259}$$

for USB and LSB receptions, respectively.

In the case of the rectangular filters and constant visibility, equations (258) and (259) are of course reduced to equations (250) and (251), since in this particular case,

$$\mathcal{V}(\omega_{LO}+\omega_I) = \mathcal{V}(\omega_{LO}-\omega_I) =\mid \mathcal{V} \mid e^{i\Phi_v},$$

and

$$\mathcal{B}_{12}^{USB}(B,\Delta\tau_g) = \mathcal{B}_{12}^{LSB}(B,\Delta\tau_g) = \frac{GB}{2}\frac{\sin(\pi B\Delta\tau_g)}{\pi B\Delta\tau_g}. \tag{260}$$

2.9.4 Fringe Amplitude and Residual Fringe Phase

The above equations (258) and (259) again show the white fringes, with bandwidth patterns in $\mid \mathcal{B}_{12} \mid$ terms and fringe patterns in cosine terms.

In general, a correlator output shows a sinusoidal fringe pattern

$$\mathcal{A}\cos\Delta\Phi,$$

near the peak of the bandwidh pattern, i.e. near the center of the coherence interval, where $\Delta\tau_g \approx 0$.

According to equations (258) and (259), theoretical expressions of the amplitude $\mathcal{A}$ and phase $\Delta\Phi$ of the fringe pattern for a continuum spectrum source are given by

$$\begin{aligned}&\mathcal{A}^U = A_0 \mid \mathcal{V}^U \mid\mid \mathcal{B}_{12}^U \mid,\\ &\Delta\Phi^U = (\omega_{LO}+\omega_I)\Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v^U - \Phi_B^U + 2\pi n^U,\end{aligned} \tag{261}$$

in the USB reception case, and

$$\mathcal{A}^L = A_0 \mid \mathcal{V}^L \mid\mid \mathcal{B}_{12}^L \mid,$$
$$\Delta\Phi^L = (\omega_{LO} - \omega_I)\Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v^L + \Phi_B^L + 2\pi n^L, \quad (262)$$

in the LSB reception case, where $2\pi n^U$ and $2\pi n^L$ are the "cycle ambiguities" inherent in phase observables.

When the fringe stopping is slightly incomplete, and therefore the phase $\Delta\Phi$ varies in time, we can see the fringe pattern as a slow sinusoidal oscillation of the correlator output (Figure 62).

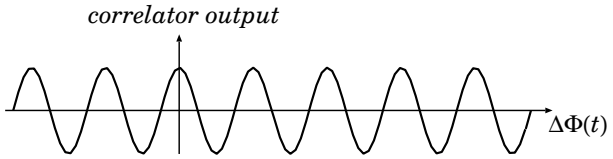

Figure 62: Slow oscillation of the correlator output with slightly incomplete fringe stopping brings the fringe pattern into view.

We call the amplitude $\mathcal{A}$ of the fringe pattern the "**fringe (or correlation) amplitude**", and the phase $\Delta\Phi$ of the fringe pattern the "**residual fringe (or correlation) phase**".

The phase $\Delta\Phi$ is often called simply as "fringe (or correlation) phase". But we prefer here the term "residual fringe phase" in order to reserve the term "fringe phase" for a quantity theoretically described by

$$\Phi = (\omega_{LO} \pm \omega_I)\tau_{g_0} + \phi_{LO1} - \phi_{LO2} - \Phi_v + \Phi_B + 2\pi n,$$

where $\pm$ correspond to the USB and LSB cases, respectively. We use the word "residual" since $\Delta\Phi$ is the residual $(\Phi - \Phi_i)$ of the "true" fringe phase Φ corrected for its theoretical prediction $\Phi_i = (\omega_{LO} \pm \omega_I)\tau_i$.

Also, in practice of interferometric observations, the fringe amplitude and residual fringe phase in the correlator outputs are frequently called as "visibility amplitude" and "visibility phase". Strictly speaking, however, one must correct the fringe amplitude and residual fringe phase for the residual delay, bandwidth pattern, and other effects, in order to obtain proper visibility amplitude $\mid \mathcal{V} \mid$ and visibility phase Φ_v as defined in equations (165) and (166).

Moreover, the residual fringe phase could be a useful observable for high precision astrometry and geodesy. This is because it contains the residual delay $\Delta\tau_g$ of the reference direction of the radio source, which could be separately obtained if one can resolve the cycle ambiguity and properly correct

the atmospheric and other disturbing effects by means of a suitable phase compensation technique. On the other hand, the visibility phase in equation (166) cannot be used for astrometry and geodesy since it is defined to be free from the residual delay of the source reference direction.

Therefore, we will henceforth distinguish the terms "fringe amplitude and residual phase" from the "visibility amplitude and phase".

2.9.5 Group Delay and Fringe Frequency

As we discussed so far, the delay tracking is aimed at reducing the slope of the phase of the cross–power spectrum of voltages to nearly zero within the observed bandwidth, thus facilitating the fringe detection in the middle of the coherence interval. Therefore, what is needed for successful delay tracking is an accurate prediction or estimation of the phase slope, i.e. frequency derivative of the phase spectrum, which can be used for precise determination of the instrumental delay.

We call the frequency derivative of the phase spectrum $\Phi(\omega, t)$ the "**group delay**" τ_G:

$$\tau_G = \frac{\partial \Phi(\omega, t)}{\partial \omega}, \tag{263}$$

where $\Phi(\omega, t)$ can be $\Phi_U(\omega, t)$ of equation (244) in case of the USB reception, or $\Phi_L(\omega, t)$ of equation (245) in case of the LSB reception. Thus, we need a good prediction or estimation of the group delay for successful delay tracking.

Usually, the most dominating term in the group delay is the geometric delay (see, for example, equation (103) in an ideally simple case).

Strictly speaking, however, equations (244) and (245) show phase terms other than the one due to the geometric delay. They may depend on the frequency and contribute to the frequency derivative. Moreover, in case of VLBI, the phase slope is affected by other effects as well, such as clock offsets among VLBI stations, atmospheric propagation delays, cable delays in signal transmission systems, and so on. These effects are often so strong that we have to take into account them for successful delay tracking. Furhermore, if some propagation delay is dispersive (i.e. frequency–dependent), then the phase slope will also depend on the frequency derivative of the dispersive delay.

On the other hand, the fringe stopping is an operation to compensate the rapid time variation of the phase of the cross–power spectrum of voltages. This is done for the purpose to integrate the product of voltages for an appropriate duration of time sufficient to detect the white fringe of a source with a good enough signal to noise ratio. Therefore, what is needed for successful fringe stopping is an accurate prediction or estimation of the time

derivative of the phase spectrum, which can be used for compensating the rapid phase shift. We call the time derivative of the phase spectrum $\Phi(\omega, t)$ the "**fringe frequency**" or "**fringe rate**", and denote it as F_r:

$$F_r = \frac{\partial \Phi(\omega, t)}{\partial t}, \tag{264}$$

where $\Phi(\omega, t)$ can be $\Phi_U(\omega, t)$ of equation (244) in case of the USB reception, or $\Phi_L(\omega, t)$ of equation (245) in case of the LSB reception, as before. Thus, we need a good prediction or estimation of the fringe frequency for successful fringe stopping.

Again, in case of VLBI, the actual fringe frequency includes time derivatives of the clock offsets, propagation delays, cable delays, and other instrmentally induced phases, besides the time derivative of the geometric delay. Some of them could have significant effects for successful fringe stopping.

For connected–element radio interferometers (CERI), an instrumental delay τ_i is usually calculated based on the geometric delay model only. The prediction is mostly accurate enough to make $\Delta\tau_{g_0}$ almost zero. Therefore, the fringe is normally found within the coherence interval, and the product of voltages can be integrated for sufficiently long time, by applying the delay tracking and fringe stopping, based on the theoretical prediction.

In VLBI, accuracies required to delay tracking and fringe stopping are much higher than in CERI, as we saw earlier. Moreover, the effects other than the geometric delay, such as clock offset and atmospheric delay, could be significant for successful delay tracking and fringe stopping.

Therefore, the theoretical prediction alone is usually not sufficient for obtaining fringes. Hence, we must first carry out a special search for estimating proper group delay and fringe frequency values, using the observed VLBI data themselves. If we succeed to accurately estimate the group delay and fringe frequency values, then we can satisfactorily perform the delay tracking and fringe stopping using these estimated values, and finally detect the correlation peak and the white fringe.

This search process is an additional labor imposed to VLBI, which is not necessary for ordinary CERI. But this "necessity" had led to the birth and the remarkable success of the geodetic VLBI as we will see later.

2.9.6 Complex Correlator

When we achieve a complete fringe stopping, the residual delay is kept always zero ($\Delta\tau_g = 0$), and, therefore, the correlator outputs given in equations (258) and (259) become constants in time. For example, in case of the USB

reception, we have

$$R^{USB}_{v_{F1}v_{F2}}(0) = A_0 \mid \mathcal{V}^U \mid\mid \mathcal{B}^U_{12} \mid \cos(\phi_{LO1} - \phi_{LO2} - \Phi^U_v - \Phi^U_B), \qquad (265)$$

where argument of cosine function (residual fringe phase) is almost constant in time.

This is not very convenient for further analysis, since, if the correlator output is just a constant, we cannot separately obtain the fringe amplitude and the residual fringe phase.

In order to get rid of this inconvenience, we use so–called complex correlators. Complex correlators are two sets of multipliers and integrators which provide two correlator outputs of the same signals, but with residual fringe phases different to each other by 90 degrees.

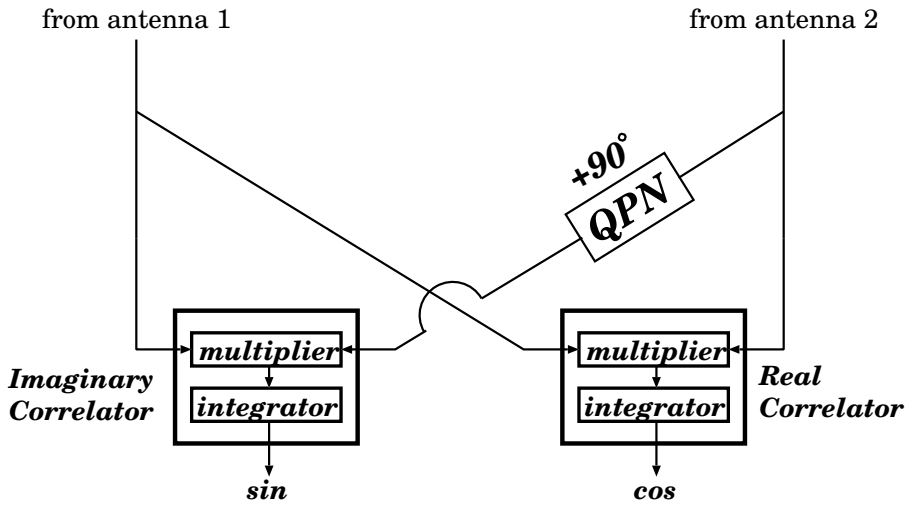

Figure 63: A complex correlator using a quadrature phase-shift network.

Figure 63 shows a design of a complex correlator which uses a quadrature phase–shift network. The quadrature phase–shift network is a special electric circuit which changes phases of all frequency components of a signal by 90 degrees.

When an output of "real correlator" $\mathcal{R}^r$, which is just an ordinary correlator we have discussed so far, shows a sinusoidal fringe pattern

$$\mathcal{R}^r = \mathcal{A} \cos \Delta\Phi,$$

where $\mathcal{A}$ and $\Delta\Phi$ are the fringe amplitude and residual fringe phase, respectively, then an output of "imaginary correlator" $\mathcal{R}^i$ with 90 degrees shifted

phase will be

$$\mathcal{R}^i = -\mathcal{A}\sin\Delta\Phi.$$

Combining the real and imaginary outputs, we can describe the output of the complex correlator in a complex form:

$$\mathcal{R} = \mathcal{R}^r + i\mathcal{R}^i = \mathcal{A}e^{-i\,\Delta\Phi}.$$

In other words, we can infer the fringe amplitude and residual fringe phase using the outputs of the real and imaginary correlators according to equations

$$\mathcal{A} = \sqrt{\mathcal{R}^{r\,2} + \mathcal{R}^{i\,2}}, \quad \text{and} \quad \Delta\Phi = -\arctan\left(\frac{\mathcal{R}^i}{\mathcal{R}^r}\right).$$

Thus, even when the fringe stopping is complete and the output of the real correlator does not vary in time as given in equation (265) for the USB reception case

$$\mathcal{R}^r = A_0 \mid \mathcal{V} \mid \mid \mathcal{B}_{12} \mid \cos(\phi_{LO1} - \phi_{LO2} - \Phi_v - \Phi_B), \tag{266}$$

we have another output from the imaginary correlator with a form

$$\mathcal{R}^i = -A_0 \mid \mathcal{V} \mid \mid \mathcal{B}_{12} \mid \sin(\phi_{LO1} - \phi_{LO2} - \Phi_v - \Phi_B). \tag{267}$$

Therefore, it is easy to separate fringe amplitude and residual fringe phase, by using outputs of these two correlators.

For a more general case when fringe is almost stopped but still a small residual delay $\Delta\tau_g$ remains, we have, from equation (258),

$$\mathcal{R}^r = \quad A_0 \mid \mathcal{V} \mid \mid \mathcal{B}_{12} \mid \cos((\omega_{LO} + \omega_I)\Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v - \Phi_B), \tag{268}$$

and

$$\mathcal{R}^i = -A_0 \mid \mathcal{V} \mid \mid \mathcal{B}_{12} \mid \sin((\omega_{LO} + \omega_I)\Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v - \Phi_B), \tag{269}$$

for real and imaginary correlator outputs in the USB reception.

Instead of the quadrature phase–shift network, we can use two LO reference signals in the frequency conversion, shifted in phase by 90 degrees.

Also, we can use multiplications of cosine and sine functions of $\omega_{LO}\tau_i$ in the fringe stopping, as we will see in discussions of VLBI correlators.

2.9.7 Projected Baseline

We discussed earlier the fringe pattern in the case when we observe a source at a direction nearly perpendicular to the interferometer baseline. When the delay tracking and fringe stopping are properly performed, however, we are in position to observe a source at any direction of the sky.

In such a case, our interferometer is equivalent to the one having a baseline length $D\cos\theta$ in Figure 64. This "effective baseline length viewed from the source" is called "**projected baseline length**". The fringe pattern is now determined by the projected baseline, as illustrated in Figure 64. Therefore, the fringe pattern now varies in time, as we track a source changing the angle θ. In particular, the fringe spacing in the sky is now given by

$$\Delta\theta_F = \frac{\lambda}{D\cos\theta}, \tag{270}$$

and also varies in time.

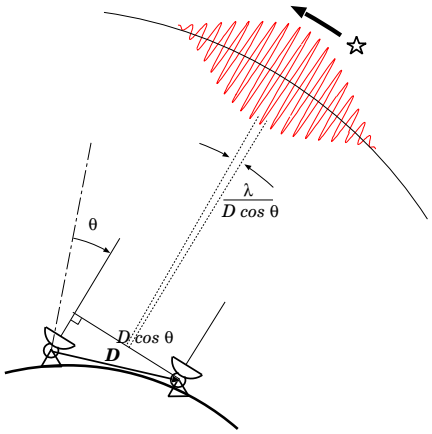

Figure 64: Projected baseline and fringe spacing at a source direction.

3 Source Structure and Correlated Flux Density

We saw earlier how a radio source structure (or, more specifically, intensity distribution) is related to an observable of the radio interferometry, the complex visibility. Using the relationship, we can infer the radio source structure by analyzing observed complex visibilities.

A number of data calibration and analysis techniques have been proposed for better recovery of source structures from usually sparce complex visibility data. They now form an extensive field of image synthesis of astronomical objects with radio interferometry, and produce high–resolution images of objects billions of light years away (see Figure 65 as an example).

However, details of the image synthesis techniques are out of scope of this lecture note. Here, we confine ourselves to a general formulation of the problem and a consideration of the effect of the finite angular size of the source on the correlator output.

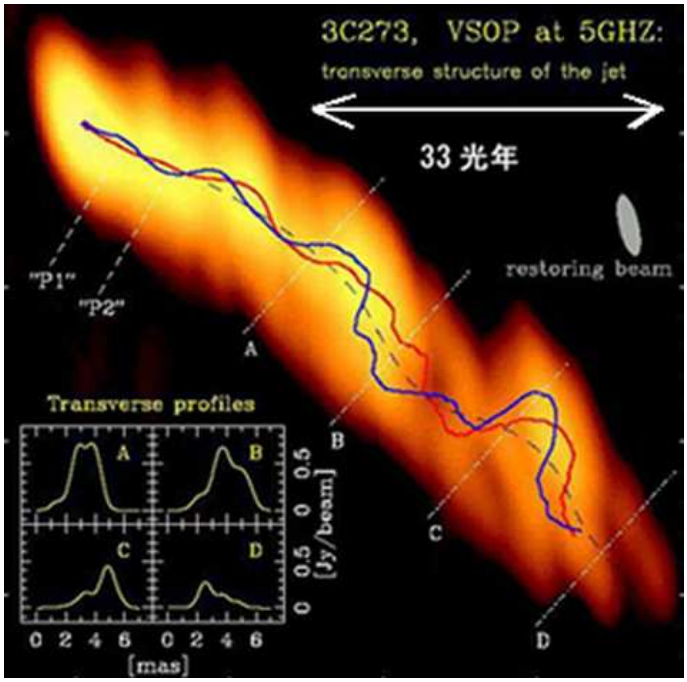

Figure 65: Double helical structure of a jet ejected from a distant quasar 3C273 revealed by a VSOP (VLBI Space Observatory Program) observation (Lobanov and Zensus, 2001; figure courtesy of ISAS/JAXA).

3.1 Basic Equations of Image Synthesis

3.1.1 Complex Visibility as an Observable

A complex correlator yields the real and imaginary correlator outputs $\mathcal{R}^r$ and $\mathcal{R}^i$, which enable us to infer the fringe amplitude and residual fringe phase for an observed radio source.

We gave in equations (268) and (269) theoretical expressions for outputs of real and imaginary correlators in a case of the USB reception:

$$\begin{aligned}\mathcal{R}^r &= \ A_0 \mid \mathcal{V} \mid \mid \mathcal{B}_{12} \mid \cos((\omega_{LO}+\omega_I)\Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v - \Phi_B),\\ \mathcal{R}^i &= -A_0 \mid \mathcal{V} \mid \mid \mathcal{B}_{12} \mid \sin((\omega_{LO}+\omega_I)\Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v - \Phi_B).\end{aligned}$$

Then, theoretical expressions for fringe amplitude and residual fringe phase are given by

$$\mathcal{A} = A_0 \mid \mathcal{V} \mid \mid \mathcal{B}_{12} \mid,$$

and

$$\Delta\Phi = (\omega_{LO}+\omega_I)\Delta\tau_g + \phi_{LO1} - \phi_{LO2} - \Phi_v - \Phi_B,$$

respectively, as we saw in equation (261).

If we achieve a high enough S/N ratio, and if we are allowed to calibrate the observed fringe amplitude and residual fringe phase for geometric mean of effective apertures A_0, the bandwidth pattern $\mid \mathcal{B}_{12} \mid e^{i\Phi_B}$, the residual delay $\Delta\tau_g$, and the difference in LO initial phases $\phi_{LO1} - \phi_{LO2}$, on the basis of suitable independent measurements and/or model estimations, we can derive the complex visibility of the observed source:

$$\mathcal{V}(\omega) = \mid \mathcal{V}(\omega) \mid e^{i\Phi_v(\omega)},$$

where we denote by ω the center frequency of the RF band, i.e., $\omega = \omega_{LO}+\omega_I$.

3.1.2 Visibility and Intensity in EW–NS Coordinate System

We showed in equation (165), that the complex visibility $\mathcal{V}(\omega)$ is related to the intensity distribution $I_\nu(\boldsymbol{\sigma})$ of an observed source as

$$\mathcal{V}(\omega) = \oint A_N(\boldsymbol{\sigma})\, I_\nu(\boldsymbol{\sigma})\, e^{-i2\pi \boldsymbol{D}_\lambda \cdot \boldsymbol{\sigma}}\, d\Omega, \tag{271}$$

where $A_N(\boldsymbol{\sigma})$ is the normalized power pattern of an interferometer, $I_\nu(\boldsymbol{\sigma})$ is the intensity distribution of the source, $\boldsymbol{D}_\lambda = \boldsymbol{D}/\lambda$ is the baseline vector $\boldsymbol{D}$ normalized by the wavelength λ of the observation, and $\boldsymbol{\sigma}$ is an offset vector of a direction $\boldsymbol{s}$ in a radio source from a reference direction $\boldsymbol{s}_0$, i.e., $\boldsymbol{\sigma} = \boldsymbol{s} - \boldsymbol{s}_0$.

The meaning of equation (271) becomes clearer if we introduce a rectangular coordinate system, whose 3-rd axis is chosen towards the reference direction $\boldsymbol{s}_0$ of the source, and 1-st and 2-nd axes are chosen in the east (right ascension) and the north (declination) directions, respectively (Figure 66).

Taking into account that $\boldsymbol{\sigma} = \boldsymbol{s} - \boldsymbol{s}_0$ is a vector difference of two unit vectors $\boldsymbol{s}$ and $\boldsymbol{s}_0$, we denote components of the offset vector $\boldsymbol{\sigma}$ in this new coordinate system as:

$$\boldsymbol{\sigma} = (\xi,\ \eta,\ \sqrt{1 - \xi^2 - \eta^2} - 1). \tag{272}$$

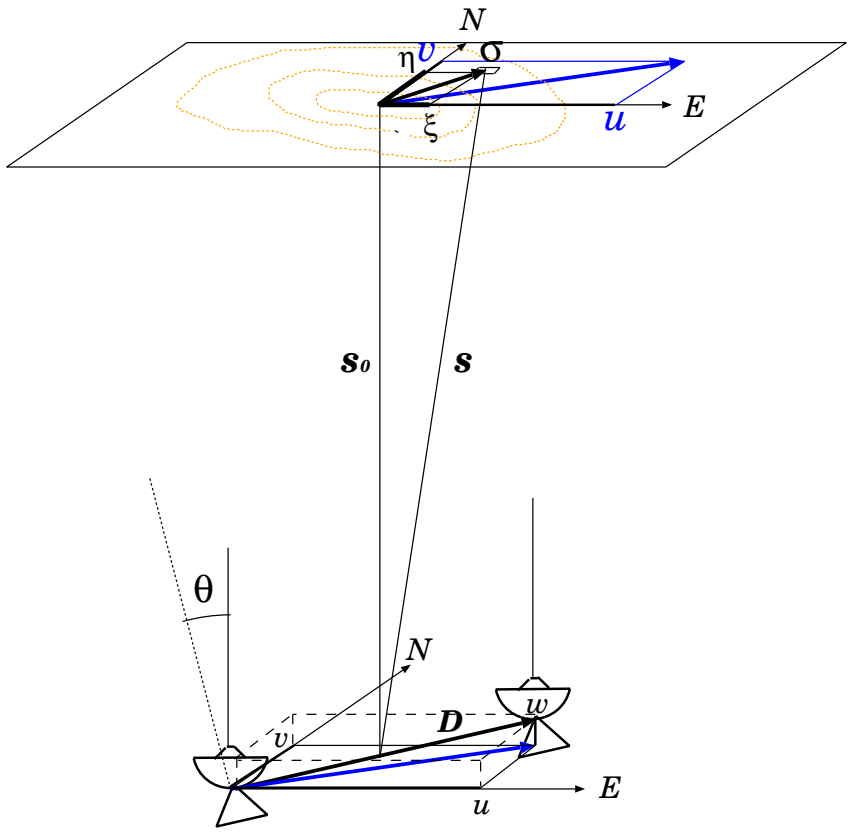

Figure 66: Source intensity distribution and projected baseline in an EW–NS plane perpendicular to the reference direction $\boldsymbol{s}_0$ of the source.

For the reason why the third component of $\boldsymbol{\sigma}$ is equal to $\sqrt{1 - \xi^2 - \eta^2} - 1$, see Figure 67.

Also, we denote components of the baseline vector $\boldsymbol{D}$ and the normalized baseline vector $\boldsymbol{D}_\lambda$ as:

$$\boldsymbol{D} = (u,\ v,\ w), \tag{273}$$

and

$$\boldsymbol{D}_\lambda = (u_\lambda,\ v_\lambda,\ w_\lambda), \tag{274}$$

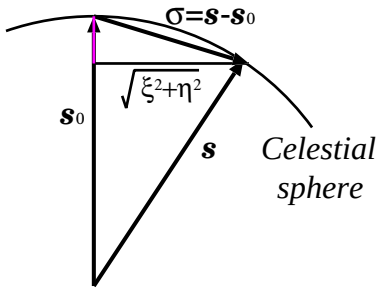

Figure 67: Third component of the offset vector $\boldsymbol{\sigma} = \boldsymbol{s} - \boldsymbol{s}_0$.

respectively, in the same coordinate system, where u_λ, v_λ, and w_λ are

$$u_\lambda = \frac{u}{\lambda}, \quad v_\lambda = \frac{v}{\lambda}, \quad \text{and} \quad w_\lambda = \frac{w}{\lambda}.$$

Note that ξ and η are taken along right ascention and declination directions, respectively, in a celestial equatorial system. Note, also, that w is related to the geometric delay τ_{g_0} of the reference direction by an equation: $w = c\tau_{g_0}$.

Then, we describe the argument of the exponential term in the complex visibility in equation (271) as:

$$2\pi \boldsymbol{D}_\lambda \cdot \boldsymbol{\sigma} = 2\pi[u_\lambda\, \xi + v_\lambda\, \eta + w_\lambda\, (\sqrt{1 - \xi^2 - \eta^2} - 1)]. \tag{275}$$

Also, the solid angle element $d\Omega$ in the same equation is given by

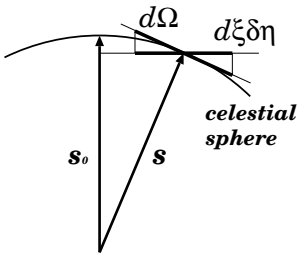

Figure 68: Solid angle element $d\Omega$ and area element $d\xi\, d\eta$ in a plane perpendicular to reference direction $\boldsymbol{s}_0$.

$$d\Omega = \frac{d\xi\, d\eta}{\sqrt{1 - \xi^2 - \eta^2}}, \tag{276}$$

since $d\xi\, d\eta$ is the projection of $d\Omega$ in a plane perpendicular to $\boldsymbol{s}_0$ (Figure 68).

Then, the complex visibility in equation (271) is given, in terms of ξ, η, u_λ, v_λ, and w_λ, by

$$\mathcal{V}(\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A_N(\xi,\,\eta) I_\nu(\xi,\,\eta) e^{-i2\pi[u_\lambda\,\xi + v_\lambda\,\eta + w_\lambda\,(\sqrt{1-\xi^2-\eta^2}-1)]} \frac{d\xi\,d\eta}{\sqrt{1-\xi^2-\eta^2}}. \tag{277}$$

This equation can be regarded as an integral equation for the intensity distribution $I_\nu(\xi,\eta)$ of a radio source with given complex visibility $\mathcal{V}(\omega)$. The purpose of the image synthesis is to get an intensity distribution $I_\nu(\xi,\eta)$ (an "image") of a radio source by solving this integral equation using observed complex visibilities.

3.1.3 Approximation of a Part of Celestial Sphere by a Tangent Plane

It is not easy to solve the integral equation (277) in its general form. However,

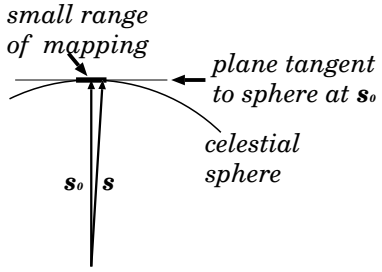

Figure 69: Celestial sphere can be approximated by a tangent plane for a small range of mapping.

if we have a narrow range of mapping, where $\xi \ll 1$ and $\eta \ll 1$, and second order terms ξ^2 and η^2 can be neglected, the equation becomes much simpler.

This narow–range condition corresponds to an approximation of a small part of the celestial sphere, where we would like to draw an image map of a radio source, by a tangent plane at the reference direction $\boldsymbol{s}_0$ (Figure 69).

Let us then examine conditions under which ξ^2 and η^2 terms can be actually neglected.

ξ^2 and η^2 terms appear in two places of equation (277), namely in

$$\frac{d\xi\,d\eta}{\sqrt{1-\xi^2-\eta^2}}, \qquad \text{and} \qquad e^{-i2\pi[u_\lambda\,\xi + v_\lambda\,\eta + w_\lambda\,(\sqrt{1-\xi^2-\eta^2}-1)]}.$$

In the first term: $\dfrac{d\xi\,d\eta}{\sqrt{1-\xi^2-\eta^2}}$, we can safely neglect ξ^2 and η^2, provided only that $|\,\xi\,| \ll 1$ and $|\,\eta\,| \ll 1$.

In the second term, which can be approximated by

$$e^{-i2\pi[u_\lambda\,\xi + v_\lambda\,\eta + w_\lambda\,(\sqrt{1-\xi^2-\eta^2}-1)]} \cong e^{-i2\pi[u_\lambda\,\xi + v_\lambda\,\eta - \frac{1}{2}\,w_\lambda\,(\xi^2+\eta^2)]}, \tag{278}$$

we must be a little careful. Here, ξ^2 and η^2 terms appear in a phase of a periodic function, for which only a residue of 2π is meaningful. Therefore, it is not always possible to neglect a term in a phase even when it is much smaller than other phase terms. A phase term can be safely neglected only when it is absolutely small as a phase (for example, much smaller than 0.1 radian, say). If we require this "$\ll 0.1$ radian" condition, then ξ^2 and η^2 in equation (278) are negligible when

$$\pi w_\lambda(\xi^2+\eta^2) \ll 0.1 \text{ radian}. \tag{279}$$

Let us now examine how this condition could be satisfied. Let us denote a

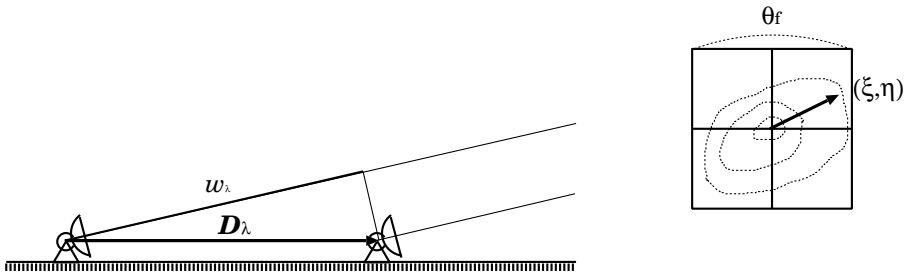

Figure 70: Length of w_λ (left) and a range of mapping (right).

full range of our intended image map (or, range of mapping) as θ_f. Then, obviously, ξ and η of a point in our source image must satisfy

$$\xi^2+\eta^2 \le (\frac{\theta_f}{2})^2,$$

(see right panel of Figure 70). Therefore, the condition in equation (279) is satisfied if

$$\pi w_\lambda(\frac{\theta_f}{2})^2 \le 0.1 \text{ radian}.$$

Adopting as a maximum value for w_λ:

$$w_{\lambda_{max}} = D_\lambda = \frac{D}{\lambda},$$

in view of left panel of Figure 70, we see that the condition in equation (279) is readily satisfied if

$$\pi \left(\frac{\theta_f}{2}\right)^2 \frac{D}{\lambda} \leq 0.1, \quad \text{and therefore} \quad \theta_f^2 \leq \frac{0.4}{\pi} \frac{\lambda}{D},$$

which is approximately equivalent to

$$\theta_f \leq \frac{1}{3} \sqrt{\frac{\lambda}{D}}. \tag{280}$$

Thus, the tangent plane approximation, where ξ^2 and η^2 can be safely neglected, is adequate, if our range of mapping θ_f satisfies equation (280).

It is still left to us to examine whether equation (280) is a realistic condition in actual mappings of radio source images.

Angular resolution θ_r of an interferometric observation is roughly given by the minimum fringe spacing:

$$\theta_r \approx \frac{\lambda}{D}.$$

Then number of grid points ("pixels") required for an image mapping is roughly proportional to $(\theta_f/\theta_r)^2$. Therefore, a load of numerical processing becomes heavier, as the range of mapping θ_f becomes wider compared with the angular resolution θ_r.

Moreover, it is usually meaningless to select a too wide range of mapping, since a radio source, which is much extended than the fringe spacing of an interferometer θ_r, tends to be "resolved out" and becomes "invisible" for the interferometer, as we will see in later discussions.

Therefore, if we select a range, which is 30 times as large as the angular resolution, as a realistic range of mapping:

$$\theta_f \approx 30 \times \theta_r \approx 30 \times \frac{\lambda}{D},$$

then, the condition of equation (280) for the tangent–plane approximation becomes

$$30 \frac{\lambda}{D} \leq \frac{1}{3} \sqrt{\frac{\lambda}{D}}, \tag{281}$$

or

$$\sqrt{\frac{\lambda}{D}} \leq \frac{1}{90}, \quad \text{and, hence,} \quad \frac{\lambda}{D} \leq \frac{1}{8100} \approx 0.^\circ 007 \approx 25 \text{ arcsec.} \tag{282}$$

As we saw in Table 2, such a condition on the fringe spacing is generally satisfied in modern interferometers, except for m–wave or cm–wave CERI arrays of relatively short baselines. In VLBI, as far as our range of mapping is $\leq 30 \times (\lambda/D)$, we can safely use the tangent–plane approximation. Figure 65 gives an example, where the angular resolution is shown as an elliptical "restoring beam" in the right edge.

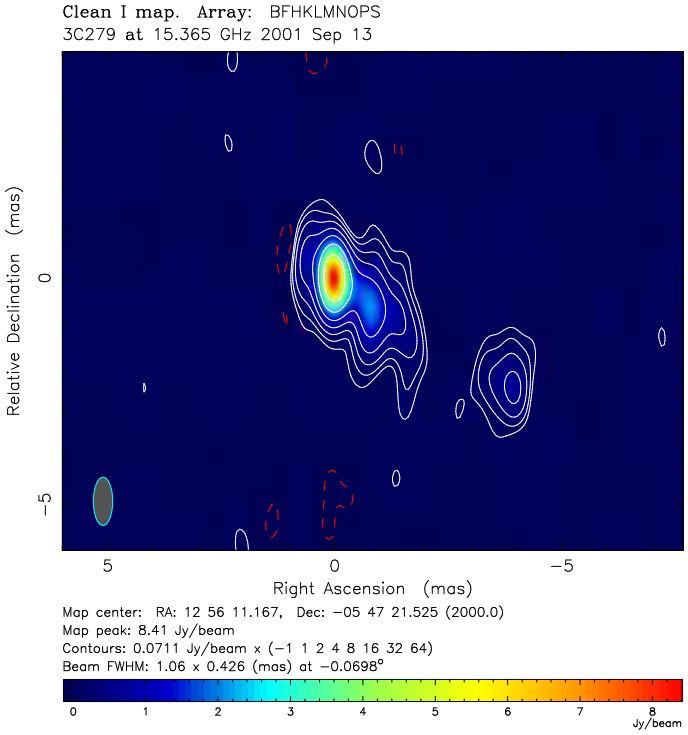

Figure 71: 15 GHz map of a quasar 3C279 observed with VLBA (Wajima and Iguchi, private communication in 2005).

Another example of a beautiful VLBI image of a quasar 3C279 observed at 15 GHz with VLBA (Wajima and Iguchi, private communication in 2005) also covers a range of mapping which is smaller than 30 times of the angular resolution (the full width half maximum (FWHM) of an interferometer beam is shown by an ellipse in the left bottom corner).

For typical CERI and VLBI, we have:

- CERI of $\lambda = 1$ cm (30 GHz) and $D = 2$ km:

$$\frac{\lambda}{D} = 5 \times 10^{-6} \text{ radian } = 1 \text{ arcsec},$$

$$\frac{1}{3}\sqrt{\frac{\lambda}{D}} = 7 \times 10^{-4} \text{ radian } = 140 \text{ arcsec.}$$

- VLBI of $\lambda = 1.35$ cm (22 GHz) and $D = 2300$ km:

$$\frac{\lambda}{D} = 6.0 \times 10^{-9} \text{ radian } = 1.2 \text{ milliarcsec (mas)},$$
$$\frac{1}{3}\sqrt{\frac{\lambda}{D}} = 2.6 \times 10^{-5} \text{ radian } = 5.3 \text{ arcsec.}$$

- VLBI of $\lambda = 2.3$ mm (129 GHz) and $D = 500$ km:

$$\frac{\lambda}{D} = 4.6 \times 10^{-9} \text{ radian } = 1.0 \text{ milliarcsec (mas)},$$
$$\frac{1}{3}\sqrt{\frac{\lambda}{D}} = 2.3 \times 10^{-5} \text{ radian } = 4.7 \text{ arcsec.}$$

In all above examples, the condition in equation (281) is well satisfied.

Important exceptions, where the tangent–plane approximation is not applicable, are wide–field mappings of some of maser sources in massive star–forming regions. They sometimes cover sky ranges of several tens arcseconds, though individual maser features, of which they consist, are compact at milliarcsecond (mas) level, and thus well detectable with intercontinental VLBIs. In such a case, the range of mapping θ_f could exceed $(1/3)\sqrt{\lambda/D}$. Therefore, one must properly take into account sphericity of the map surface in such wide–field mappings.

3.1.4 Interferometer is a Fourier Transformer

Under the tangent plane approximation, equation (277) is reduced to

$$\mathcal{V}(\omega) = \mathcal{V}(u_\lambda, v_\lambda) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A_N(\xi,\,\eta) I_\nu(\xi,\,\eta) e^{-i2\pi(u_\lambda\,\xi + v_\lambda\,\eta)} d\xi\, d\eta, \qquad (283)$$

where we reexpressed $\mathcal{V}(\omega)$ as $\mathcal{V}(u_\lambda, v_\lambda)$, stressing dependence of the visibility on "normalized projected baseline components" u_λ and v_λ.

Then, it is evident that the complex visibility $\mathcal{V}(u_\lambda, v_\lambda)$, as given in equation (283), is a two–dimensional Fourier transform, or a "spectrum", of the source intensity (or brightness) distribution $I_\nu(\xi, \eta)$ multiplyed by $A_N(\xi,\,\eta)$ at "spatial frequency" values u_λ and v_λ.

Consequently, we can regard a radio interferometer as a "Fourier Transformation Device", which yields, at the correlator output, a two–dimensional Fourier component or "spatial frequency spectrum" of the source intensity distribution times $A_N(\xi, \eta)$ on the sky, at spatial frequencies corresponding to two components of the normalized projected baseline, u_λ and v_λ. The Fourier transformation relation described in equation (283) is called "van Cittert–Zernike Theorem".

Naturally, the source intensity distribution is restored from the measured visibilities by the inverse Fourier transformation:

$$A_N(\xi, \eta) I_\nu(\xi, \eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{V}(u_\lambda, v_\lambda) e^{i2\pi(u_\lambda \xi + v_\lambda \eta)} \, du_\lambda dv_\lambda, \qquad (284)$$

and then by correcting the normalized power pattern of the interferometer $A_N(\xi, \eta)$. This is the basis of the astrophysical imaging of the radio source structure by means of radio interferometry under the tangent plane approximation. This is important also for geodetic and astrometric VLBI, when we calibrate the effect of the source structure on the measured delay.

Of course, it is impossible to perform a complete inverse Fourier transformation, unless a continuous distribution of the visisbility is available over the whole spatial frequency domain, the "uv–plane". Nevertheless, a better restoration is achieved, with suitable image processing software, when the visibilities are sampled at many u_λ, v_λ points, which nearly homogeneously cover an area of the uv–plane contained within a radius D_{max}/λ from the origin, where D_{max} is the maximum baseline length in an iterferometer array. It is therefore important to achieve a good coverage of the uv–plane in designing an interferometer array aimed at astrophysical imaging of radio sources.

3.1.5 Fringe Spacing as Angular Resolution of Interferometer

Our "spatial frequency spectrum" of the source intensity distribution, i.e. the complex visibility, $\mathcal{V}(u_\lambda, v_\lambda)$ lacks spatial frequency components higher than D_{max}/λ, simply because the finite size of our interferometer does not allow us to obtain the complex visibility in the higher range. This means that the inverse Fourier transformation is incapable of restoring structures smaller than λ/D_{max}.

This is the reason why the angular resolution of an interferometer array is determined by its minimum fringe spacing λ/D_{max}.

Note that, although we formally set the range of integration in equation (284) from $-\infty$ to ∞, the actual range is limited by the maximum uv–length

D_{max}/λ.

3.2 uv–Coverage

The Earth's diurnal rotation changes the direction and length of a projected baseline of a ground–based interferometer in the uv–plane, and thus helps us to sample visibilities at different u_λ, v_λ points. These visibilities, though sampled at different times, can well be used for restoring images of sources whose structures do not appreciably change within a day.

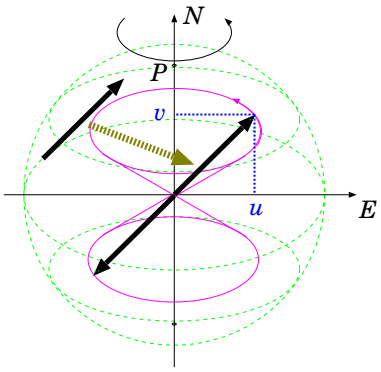

Figure 72: If we place an end of a baseline vector at the center of the rotating Earth, by a parallel translation, a tip of the vector draws an ellipse, in general, viewed from a radio source. EW and NS components of the projected baseline are the u and v components. Since we have two vectors of equal lengths and opposite directions which connect two stations, we have two point–symmetric elliptical trajectories.

Since the interferometer baseline rotates around a nearly fixed spin axis of the Earth, the trajectory of the projected baseline is a circle, if the observed source is located at the celestial pole. The trajectory is a linear oscillation when the source is at the equator, and is an ellipse, for all other directions in between the celestial equator and the 2 celestial poles. For two stations 1 and 2 on the surface of the Earth, we have two baseline vectors with opposite directions $\boldsymbol{D}_{12}$ and $\boldsymbol{D}_{21} = -\boldsymbol{D}_{12}$ connecting these two stations. Therefore, in general, we have two point–symmetric arcs of ellipses on the uv–plane for a pair of stations (Figure 72).

Configurations of radio interferometer arrays are designed to achieve good coverages of the uv–plane, from the trajectories of projected baselines formed by their antennas.

3.2.1 uv–Trajectories

Figure 73 shows geometry of the diurnal rotation of a baseline vector $\boldsymbol{D}$.

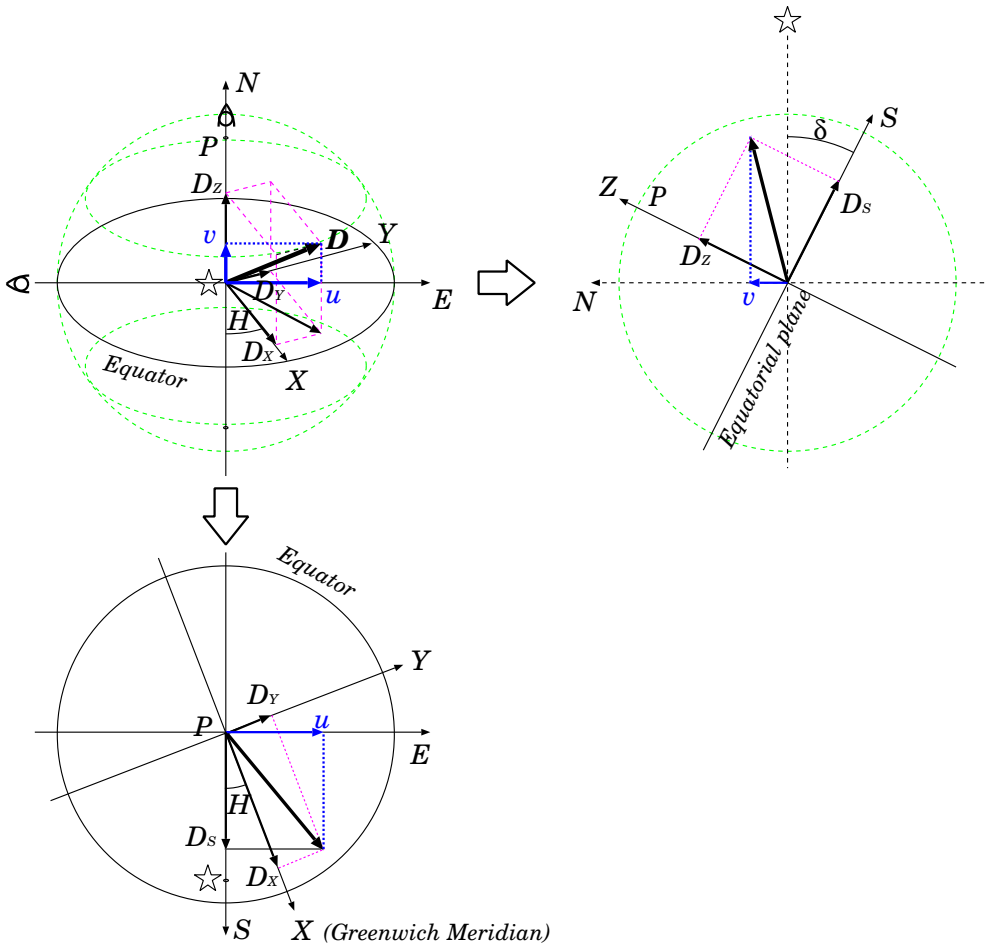

Figure 73: Geometrical relationship between D_X, D_Y, and D_Z components of a baseline vector $\boldsymbol{D}$ and u, v coordinates. Here, X, Y, Z coordinate system is an Earth–fixed right–handed Cartesian system with Z axis towards the North Pole P and X axis towards Greenwich Meridian. H and δ are Greenwich hour angle and declination of an observed radio source. The Earth–fixed baseline vector is shown as viewed from the radio source (top, left), viewed from west (top, right), and viewed from the North Pole (bottom).

We use an Earth–fixed Cartesian X, Y, Z coordinate system with Z axis directed towards North Pole, X axis directed towards Greenwich Meridian, and Y axis which completes a right–handed system. Components of the baseline vector $\boldsymbol{D}$ in this system are denoted by D_X, D_Y, and D_Z. Direction of an observed radio source is given by the declination δ and Greenwich hour

angle H of the source with respect to this Earth–fixed coordinate system.

First, let us look at the baseline vector from a direction of the North Pole (bottom–left panel of Figure 73). We see two sets of rectangular axes X, Y and S, E in the equatorial plane, which is now perpendicular to our line of sight. S axis here is chosen along a plane containing both the North Pole axis and the radio source direction, which we call the "PS–plane". We then consider S and E components of the baseline vector $\boldsymbol{D}$ in the equatorial plane. E component is just equal to the u component, i.e. the EW–component of the baseline vector viewed from the radio source (top–left panel of Figure 73). As for S component of the baseline $\boldsymbol{D}$, we denote it as D_S. Since S, E and X, Y axes are inclined to each other by the Greenwich hour angle H, the u and D_S components and D_X and D_Y components of the baseline vector $\boldsymbol{D}$ are related to each other by equations:

$$u = D_X \sin H + D_Y \cos H, \tag{285}$$

and

$$D_S = D_X \cos H - D_Y \sin H. \tag{286}$$

Next, when we look at the baseline vector $\boldsymbol{D}$ along the E axis from the west side (top–right panel of Figure 73), we see equatorial S axis and polar Z axis in the PS–plane, which is now perpendicular to the line of sight. The radio source direction is offset from the equatorial plane by the declination δ. Therefore, the component of the baseline vector $\boldsymbol{D}$ in an axis perpendicular to the source direction, which is nothing but the NS–component of the baseline vector viewed from the radio source (top–left panel of Figure 73), i.e. the v component, is related to the D_S and D_Z components by an equation:

$$v = -D_S \sin\delta + D_z \cos\delta. \tag{287}$$

Combining equations (286) and (287), we obtain

$$v = -(D_X \cos H - D_Y \sin H) \sin\delta + D_Z \cos\delta, \tag{288}$$

for v component. Equations (285) and (288) allow us to calculate a uv–trajectory of a radio source with varying Greenwich hour angle H.

From equations (285) and (288), we obtain an equation of ellipse:

$$u^2 + \frac{(v - D_Z \cos\delta)^2}{\sin^2\delta} = D_X^2 + D_Y^2, \tag{289}$$

which clearly shows that uv–trajectories are really ellipses.

In view of the point symmetry property mentioned above, a set of u and v:

$$\begin{aligned} u &= -D_X \sin H - D_Y \cos H, \\ v &= (D_X \cos H - D_Y \sin H) \sin \delta - D_Z \cos \delta, \end{aligned} \tag{290}$$

which satisfies an equation of another ellipse:

$$u^2 + \frac{(v + D_Z \cos \delta)^2}{\sin^2 \delta} = D_X^2 + D_Y^2, \tag{291}$$

is also a valid uv–trajectory of the same pair of stations.

When we calculate actual uv–trajectory of a radio source, we must take into account a fact that the source can be observed by an interferometer baseline only when the source is "mutually visible" from two stations of the baseline, i.e. when the source is within lower and upper elevation limits of radio telescopes at the two stations. This condition is roughly formulated as follows.

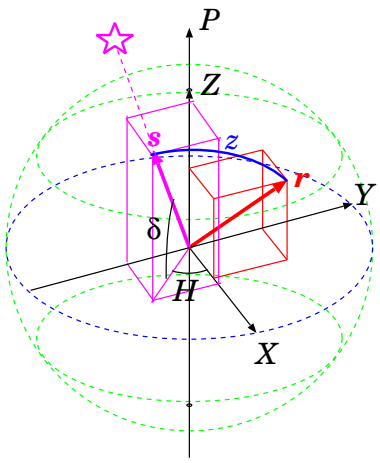

Figure 74: Zenith distance z of a radio source direction $\boldsymbol{s}$ at a station in a direction $\boldsymbol{r}$. Also shown is an Earth–fixed Cartesian coordinate system with Z–axis towards the North Pole P and X–axis towards Greenwich Meridian. H and δ are Greenwich hour angle and declination of the source, respectively.

If X, Y, Z coordinates of a station is $\boldsymbol{R} = (R_X, R_Y, R_Z)$, and if we approximate the Earth by a sphere, then a unit vector $\boldsymbol{r}$ oriented towards zenith direction of the station is roughly given by

$$\boldsymbol{r} = \frac{\boldsymbol{R}}{R},$$

where

$$R = \sqrt{R_X^2 + R_Y^2 + R_Z^2}.$$

On the other hand, components of a unit vector $\boldsymbol{s}$ towards a radio source are given in the X, Y, Z coordinates by

$$\boldsymbol{s} = (\cos\delta\cos H,\ -\cos\delta\sin H,\ \sin\delta),$$

(see Figure 74), where, again, H is the Greenwich hour angle, and δ is the declination of the source.

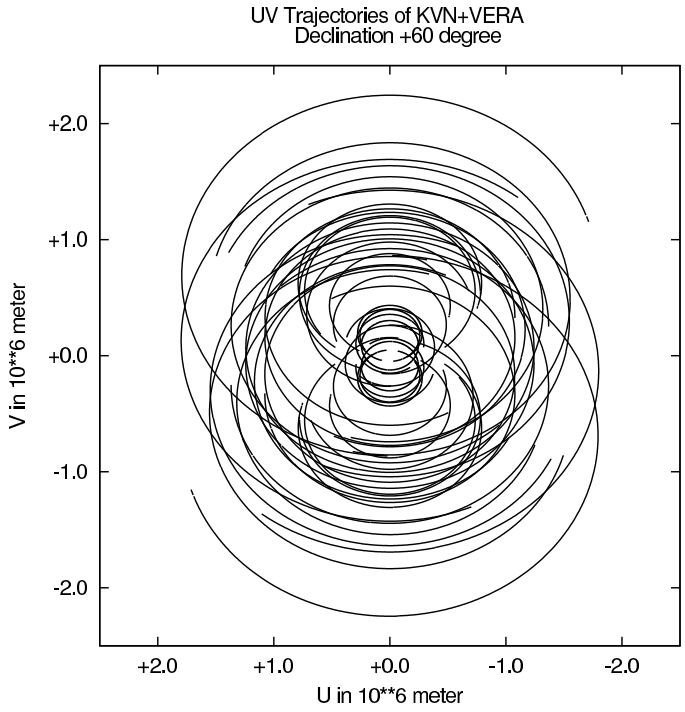

Figure 75: *uv*–trajectories of a combined KVN and VERA arrays, for a source at $\delta = +60°$.

Therefore, cosine of the zenith distance z, i.e. the angle between the zenith direction $\boldsymbol{r}$ and the source direction $\boldsymbol{s}$, is approximately given by

$$\cos z = \boldsymbol{r}\cdot\boldsymbol{s} = \frac{R_X\cos\delta\cos H - R_Y\cos\delta\sin H + R_Z\sin\delta}{R}. \qquad (292)$$

Thus, if upper and lower elevation limits of a radio telescope at a station are E_u and E_l, respectively, then the radio source is visible with the telescope when

$$\sin E_u \le \cos z \le \sin E_l. \qquad (293)$$

Actual uv–trajectories can be calculated in terms of equations (285), (288), and (290), when the condition of equation (293) is satisfied at both ends of baselines.

Figure 75 shows an example of the uv–coverage for the combined VERA and KVN VLBI arrays thus calculated, for a source located at +60° declination.

3.2.2 Synthesized Beams

Let us consider a point source with flux density $\mathcal{S}_\nu$ located at the reference direction $\boldsymbol{s}_0$ ($\xi = 0$ and $\eta = 0$). In this case, the intensity distribution, which we call here the "true intensity" and denote as $I_\nu^{true}(\xi, \eta)$, in order to distinguish it from "synthesized" one to be introduced below, is given by

$$I_\nu^{true}(\xi, \eta) = \mathcal{S}_\nu \, \delta(\xi) \, \delta(\eta), \tag{294}$$

where $\delta(x)$ is the delta function.

For simplicity, we assume that the normalized power pattern of the interferometer is unity, i.e. $A_N(\xi, \eta) = 1$, everywhere in our range of mapping. Then, equation (283) gives the complex visibility which is constant at every u_λ, v_λ point:

$$\mathcal{V}(u_\lambda, v_\lambda) = \mathcal{S}_\nu = \text{const.} \tag{295}$$

In an actual interferometer array, we can sample the complex visibility only along uv–trajectories. Therefore, the "synthesized intensity distribution" calculated by equation (284):

$$I_\nu^{synt}(\xi, \eta) = \mathcal{S}_\nu \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i2\pi(u_\lambda \xi + v_\lambda \eta)} du \, dv \Big|_{along\ uv} , \tag{296}$$

does not reproduce the point but yields some extended distribution. This "response of a synthesis interferometer array to a point source" is called the "synthesized beam" or the "dirty beam" of the array. This is analogous to the beam formation of a single dish antenna with a strange "aperture illumination" along elliptical arcs.

If we approximate the two–dimensional integration by a summation on meshes, equation (296) is reduced to

$$\begin{aligned} I_\nu^{synt}(\xi, \eta) &= \mathcal{S}_\nu \sum_u \sum_v e^{i2\pi(u_\lambda \xi + v_\lambda \eta)} \Delta u_\lambda \, \Delta v_\lambda \Big|_{along\ uv} \\ &= \mathcal{S}_\nu \sum_u \sum_v [e^{i2\pi(u_\lambda \xi + v_\lambda \eta)} + e^{-i2\pi(u_\lambda \xi + v_\lambda \eta)}] \Delta u_\lambda \, \Delta v_\lambda \Big|_{along\ uv1} \\ &= 2\mathcal{S}_\nu \sum_u \sum_v \cos(2\pi[u_\lambda \xi + v_\lambda \eta]) \Delta u_\lambda \, \Delta v_\lambda \Big|_{along\ uv1} , \end{aligned} \tag{297}$$

where "*along uv*" means summation along *uv*–trajectories composed of point–symmetric elliptical arcs given in both equations (289) and (291), whereas "*along uv*1" means summation along *uv*–trajectories given by equation (289) only.

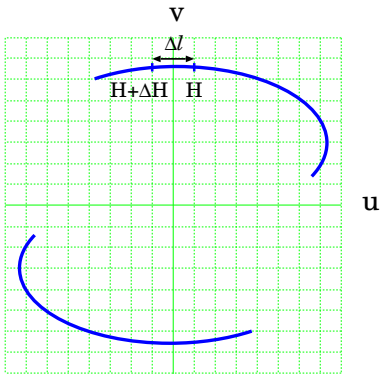

Figure 76: Integration along elliptical *uv*–trajectories.

We can further replace the two–dimensional summation on the *uv*–plane by a one–dimensional summation along the elliptical *uv*–trajectories with equal intervals of Greenwich hour angle ΔH (Figure 76). Since an interval in Greenwich hour angle ΔH corresponds to a length interval Δl in the *uv*–plane

$$\Delta l = \sqrt{\left(\frac{du_\lambda}{dH}\right)^2 + \left(\frac{dv_\lambda}{dH}\right)^2}\,\Delta H,$$

we obtain from equation (297)

$$I_\nu^{synt}(\xi,\,\eta) \propto \sum_H \cos(2\pi[u_\lambda(H)\xi + v_\lambda(H)\eta])\,\sqrt{\left(\frac{du_\lambda}{dH}\right)^2 + \left(\frac{dv_\lambda}{dH}\right)^2}\,\Delta H\Big|_{along\ uv1}\,. \tag{298}$$

Now, u_λ, v_λ and their derivatives with respect to H in equation (298) are readily calculated for a radio source with declination δ from equations (285) and (288), namely we have

$$u_\lambda = \frac{D_X \sin H + D_Y \cos H}{\lambda}, \tag{299}$$

$$v_\lambda = -\frac{D_X \cos H - D_Y \sin H}{\lambda}\sin\delta + \frac{D_Z}{\lambda}\cos\delta, \tag{300}$$

and

$$\frac{du_\lambda}{dH} = \frac{D_X \cos H - D_Y \sin H}{\lambda}, \tag{301}$$

$$\frac{dv_\lambda}{dH} = \frac{D_X \sin H + D_Y \cos H}{\lambda} \sin\delta. \tag{302}$$

Calculating these quantities with varying H while the source is mutually visible at two ends of baselines, inserting them into equation (298), and dividing the results by a maximum value, we obtain the synthesized beam normalized by its maximum value:

$$\frac{I_\nu^{synt}(\xi,\,\eta)}{I_\nu^{synt}{}_{max}}.$$

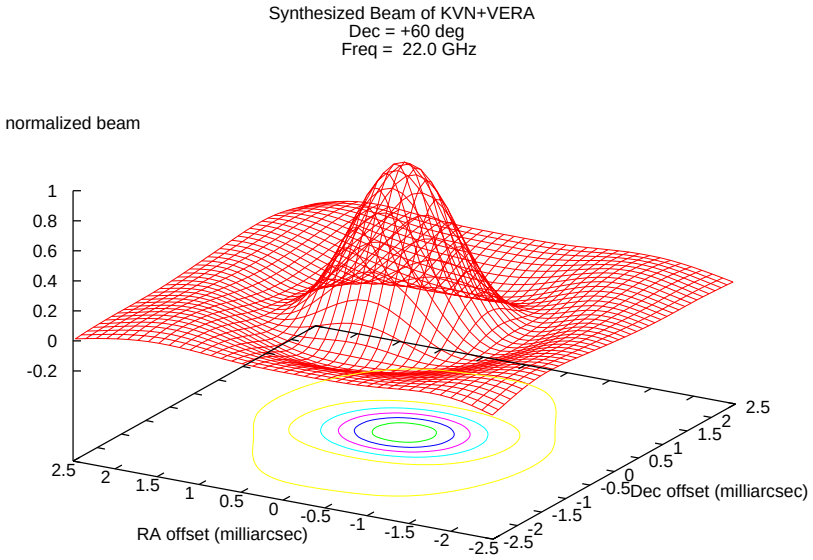

Figure 77: Normalized synthesized beam of the combined KVN and VERA VLBI arrays for a radio source at +60° declination with observing frequency 22 GHz.

Figure 77 shows an example of the normalized synthesized beam for the combined KVN and VERA VLBI arrays for a radio source located at +60° declination observed at 22 GHz. It is clear from this Figure that beam width of this synthesized beam is around 1 mas which corresponds to the minimum fringe spacing λ/D_{max} with $\lambda = 1.4$ cm (22GHz) and $D_{max} \cong 2300$ km. This again shows that the angular resolution of an interferometer is determined by the minimum fringe spacing.

3.3 Correlated Flux Density of a Source with Gaussian Intensity (Brightness) Distribution

How is the correlated flux density (or visibility amplitude) related to the size of an observed radio source?

Let us consider this problem for a simple case, when the source has a circular Gaussian intensity (brightness) distribution (Figure 78):

$$I_\nu(\xi,\,\eta) = \frac{\mathcal{S}_{T_\nu}}{\pi\Theta_s^2} e^{-\frac{\xi^2+\eta^2}{\Theta_s^2}}, \tag{303}$$

where $\mathcal{S}_{T_\nu}$ is the total flux density (note that $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} e^{-\frac{\xi^2+\eta^2}{\Theta_s^2}} d\xi\, d\eta = \pi\Theta_s^2$),

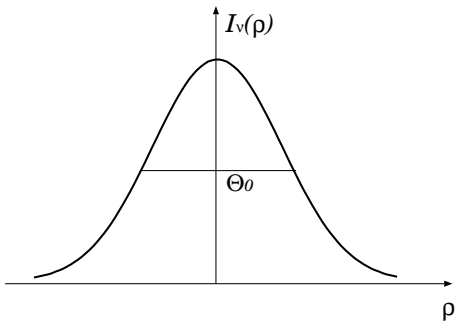

Figure 78: Circular Gaussian intensity distribution $I_\nu(\rho)$ with $\rho = \sqrt{\xi^2+\eta^2}$.

and Θ_s is a parameter characterizing the source size. This parameter Θ_s is related to the half–power width of the Gaussian distribution Θ_0 by a formula:

$$\Theta_s = \frac{\Theta_0}{2\sqrt{\ln 2}} \cong 0.60\Theta_0. \tag{304}$$

Then, in view of equation (283), the complex visibility of the source is given by

$$\begin{aligned}\mathcal{V}(u_\lambda,\, v_\lambda) &= \frac{\mathcal{S}_{T_\nu}}{\pi\,\Theta_s^2} \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} e^{-\frac{\xi^2+\eta^2}{\Theta_s^2}} e^{-i2\pi(u_\lambda\,\xi + v_\lambda\,\eta)}\, d\xi\, d\eta \\ &= \mathcal{S}_{T_\nu}\, e^{-\pi^2\,\Theta_s^2\,(u_\lambda^2+v_\lambda^2)} \;\; = \;\; \mathcal{S}_{T_\nu}\, e^{-\left[\pi\,\Theta_s\left(\frac{D\,\cos\theta}{\lambda}\right)\right]^2}\end{aligned} \tag{305}$$

(see Figure 64 for geometry of the projected baseline), where we took $A_N(\xi,\,\eta) \cong 1$, for simplicity, assuming that the source is observed at beam centers of an-

tennas. In deriving equation (305), we used an integration formula:

$$\int_{-\infty}^{\infty} e^{-x^2 - i\,a\,x}\,dx = \sqrt{\pi}\,e^{-\frac{a^2}{4}}.$$

Thus, for a source with the circular Gaussian intensity distribution, we have

- correlated flux density $= \mid \mathcal{V}(u_\lambda,\, v_\lambda) \mid = \mathcal{S}_{T_\nu}\, e^{-\left[\pi\,\Theta_s\,\left(\frac{D\,\cos\theta}{\lambda}\right)\right]^2}$,
- visibility phase $= 0$.

This means that the correlated flux density $\mid \mathcal{V}(u_\lambda,\, v_\lambda \mid$ is reduced to the half of the total flux density $\mathcal{S}_{T_\nu}$ when

$$\left[\pi\Theta_s(\frac{D\cos\theta}{\lambda})\right]^2 = \ln 2, \tag{306}$$

or, in view of equation (304), when the half–power width of the source distribution Θ_0 is equal to

$$\Theta_0 = \frac{2\ln 2}{\pi}\frac{\lambda}{D\cos\theta} \cong 0.44\frac{\lambda}{D\cos\theta}. \tag{307}$$

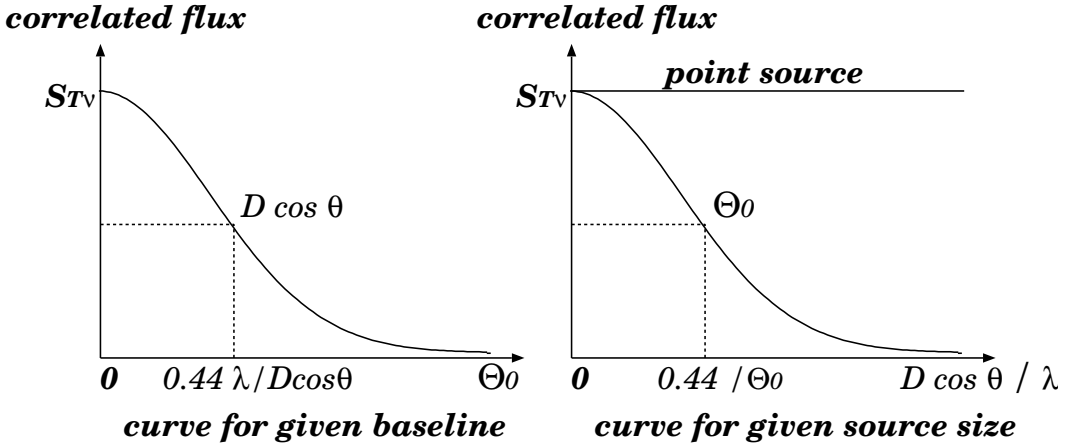

Figure 79: Correlated flux density of a source with Gaussian intensity distribution, as a function of the source size (left), and as a function of the projected baseline length (right).

Figure 79 shows the correlated flux density of the source, with a circular Gaussian intensity distribution, as a function of the half–power width Θ_0 of the source intensity distribution, assuming a fixed baseline length D (left panel); and as a function of the projected baseline length normalized by the

wave length, $D\cos\theta/\lambda$, assuming a fixed source size (right panel), according to equation (305). This simple example shows that the source appears weak to an interferometer, or "resolved out", when its size becomes comparable with or larger than the fringe spacing $\lambda/(D\cos\theta)$. In other words, interferometers are sensitive to the sources with angular sizes comparable to or smaller than their fringe spacings.

It is usual practice in VLBI to infer the size of a source by measuring its correlated flux densities at different projected baseline lengths, and fitting the results to a Gaussian model, as shown in the right panel of Figure 79. Figure 80 shows an example (Wajima and Iguchi, private communication in 2005).

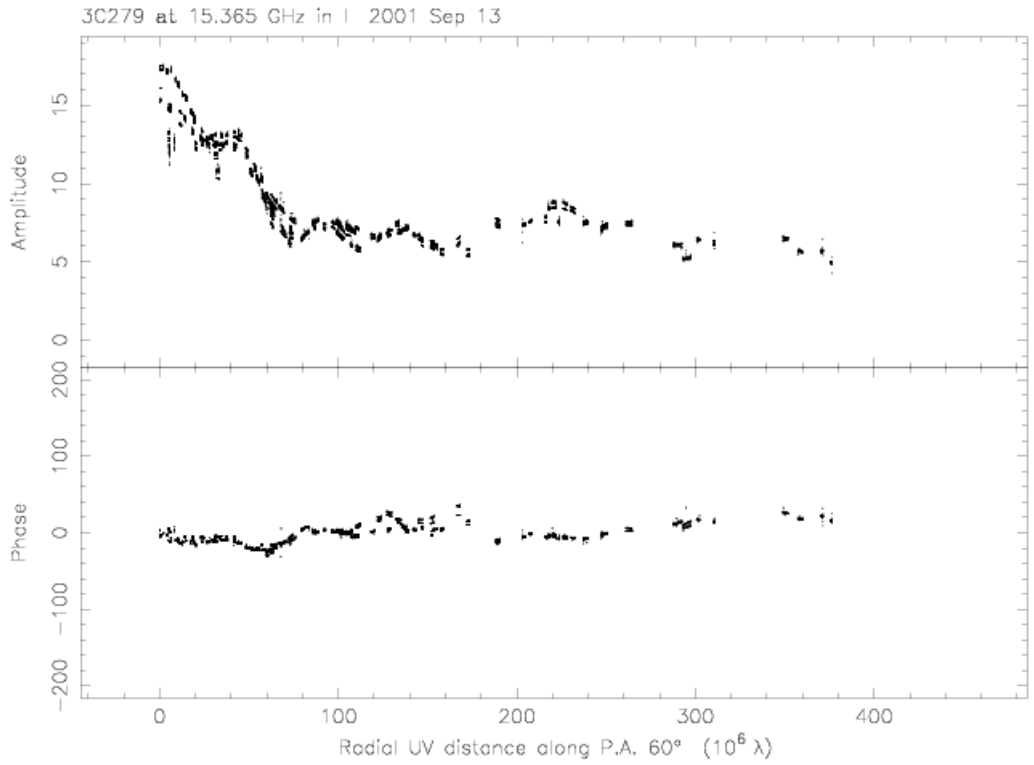

Figure 80: Visibility amplitude (top) and phase (bottom) versus uv–distance along a position angle (PA) 60° plot of a quasar 3C279 observed with VLBA at 15 GHz (Wajima and Iguchi, private communication in 2005). Visibility amplitude (top) decreases with uv–distance (projected baseline length) in a range from 0 to 70×10^6 λ, similarly to the curve in the right panel of Figure 79. However, visibility amplitude becomes almost flat in uv–distance range from 70×10^6 to 400×10^6 λ, showing existence of a compact component which is not resolved even on the longest baseline 400×10^6 $\lambda \simeq 8000$ km. This Figure can be compared with an image map of the same source shown in Figure 71.

3.4 What Can VLBI Observe?

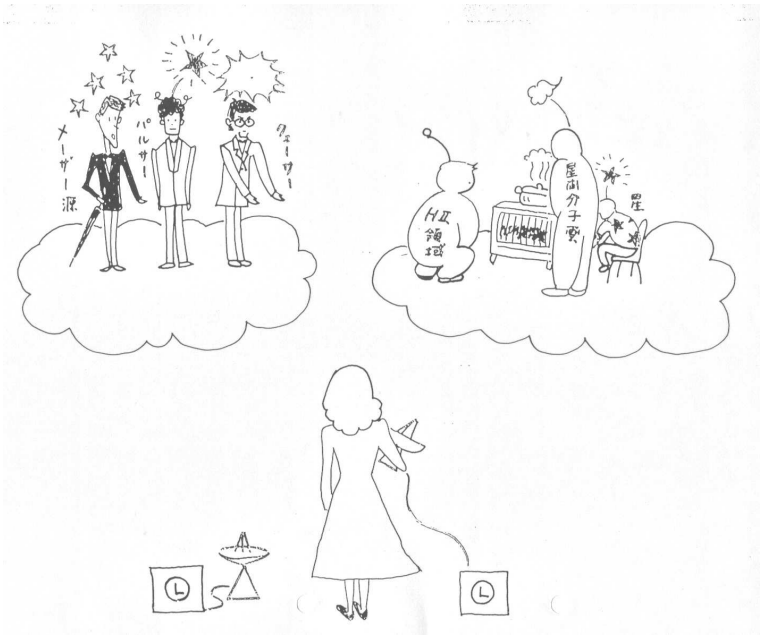

Figure 81: VLBI can observe very bright, compact objects only.

The above discussions show that a VLBI (or an interferometer, in general), with baseline length D and observing wavelength λ, can detect only compact sources with characteristic angular diameters Θ_0 comparable to or smaller than λ/D, i.e.,

$$\Theta_0 \leq \frac{\lambda}{D}.$$

If the brightness temperature of a source is T_B, then the intensity I_ν and the flux density $\mathcal{S}_\nu$ are given by:

$$I_\nu = \frac{2kT_B}{\lambda^2}, \quad \text{and} \quad \mathcal{S}_\nu = \frac{2kT_B}{\lambda^2}\frac{\pi\Theta_0^2}{4} \leq \frac{\pi k T_B}{2D^2},$$

where $k = 1.381 \times 10^{-23}$ JK^{-1} is the Boltzmann constant. Note that the last expression does not depend on the wavelength.

Now, if the minimum detectable flux density by VLBI is $\mathcal{S}_{\nu min}$, we can impose a lower limit to the brightness temperature, T_{Bmin}, for the detectable source:

$$T_{Bmin} \cong \frac{2D^2}{\pi k}\mathcal{S}_{\nu min}. \tag{308}$$

If $D = 2300$ km and $\mathcal{S}_{\nu min} = 0.02$ Jy, then $T_{Bmin} \cong 5 \times 10^7$ K (!).

Such a high lower limit of the brightness temperature precludes detection of almost all thermal sources, such as stars and molecular clouds with VLBI, leaving only very bright, compact non–thermal sources, e.g., AGNs, masers, and pulsars (see Figure 81), to be detected.

4 Sensitivity of Radio Interferometer

Now let us consider how we can estimate sensitivity of a radio interferometer, i.e. signal–to–noise ratio of its correlator output.

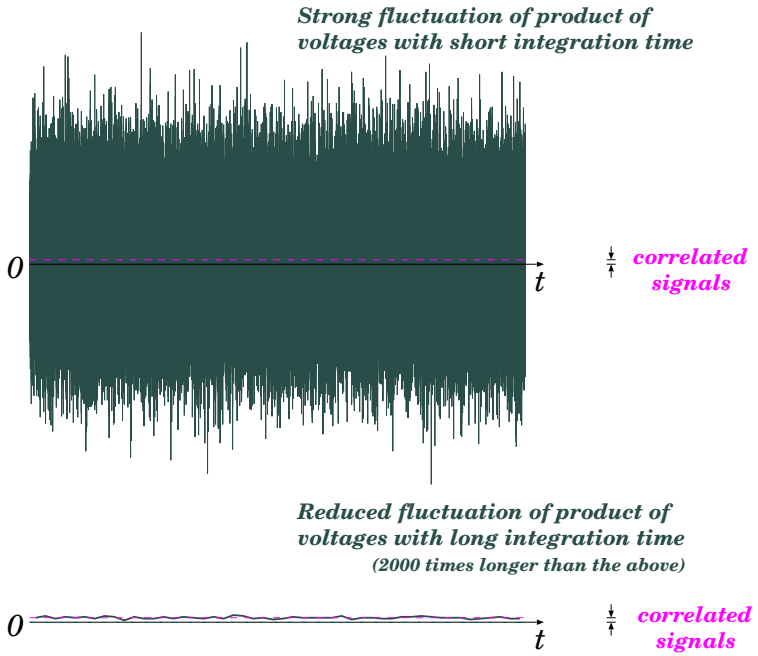

Figure 82: Simulated temporal behaviors of correlator outputs as time averaged products of received voltages with short (top) and long (bottom) integration (averaging) times. The voltages are assumed to consist of a weak common signal and strong independent noises. Time averaged product of the independent noises strongly fluctuates in time t when the integration time is short but tends to zero as the integration time increases, while that of the signal components tends to a positive constant (correlated signals). Thus, signal–to–noise ratio of an interferometer is improved with integration time.

When we discussed a realistic interferometer, we had taken into account only signal components of received voltages, ignoring any system noise contri-

bution. However, now we have to remember that the system noise is inherent in every radio telescope and usually much stronger than the signal from an astronomical radio source.

Figure 82 shows results of a simulation analogous to what we considered in subsection 1.2.4 where we discussed the ergodicity. In order to model a time series of a voltage received by an element antenna of an interferometer, we successively generated a series of superposed two independent Gaussian random numbers with substantially different dispersions. They stand for a small signal and a big system noise contained in the voltage. Then we calculated a "correlator output" by time averaging a product of two such voltages, regarding them as "received at two antennas". For simplicity, we assumed that signals are identical and noises are independent but with equal dispersion in the two received voltages. Note that the signals could be identical in an interferometer provided that the delay tracking and fringe stopping are perfectly performed, element antennas and receiving systems are identical, and observed source is well compact compared with the fringe spacing.

Upper and lower panels of Figure 82 show temporal behaviors of the simulated correlator output when integration (averaging) time is short (top) and long (bottom), respectively. The integration time in the lower panel is 2000 times longer than that in the upper panel.

While the integration time is short, the correlator output violently fluctuates due to the strong noises burying any sign of small signals (top). As integration time increases, a time–averaged product of common signals tends to a positive constant, i.e. the expected correlation of source signals, which we call here the "correlated signals". On the other hand, a time–averaged product of independent noises naturaly tends to zero. Therefore, when the integration time is long enough to suppress the noise fluctuation, we certainly see the correlated signals (bottom).

We define the signal–to–noise ratio of a correlator output of a radio interferometer as a quantity showing how well we can detect a source through integration (time–averaging) of a product of received voltages.

Specifically, we define it as a ratio between the correlated signals and a standard deviation (level of fluctuation) of a correlator output, i.e. a time–averaged product of received voltages.

In the followings, we will derive a theoretical expression for the signal–to–noise ratio S/N on the basis of a statistical theory of the standard deviation of the time–averaged product of random processes.

4.1 Statistical Model of a Correlator Output

4.1.1 Signal + Noise Inputs to a Correlator

Let us consider that received voltages $U_1(t)$ and $U_2(t)$ from two antennas, which are fed to input points of a correlator, consist of signals $V_1(t)$, $V_2(t)$

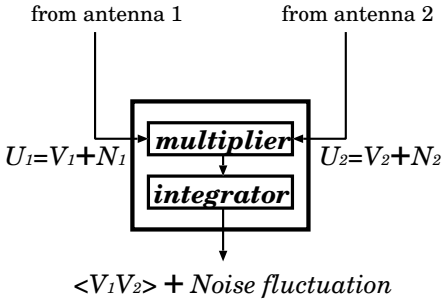

Figure 83: Input voltages to a correlator consist of signals and noises.

and noises $N_1(t)$, $N_2(t)$ (see Figure 83):

$$\begin{aligned} U_1(t) &= V_1(t) + N_1(t), \\ U_2(t) &= U_2(t) + N_2(t). \end{aligned} \tag{309}$$

We assume that the delay tracking and fringe stopping were already successfully applied to the voltages $U_1(t)$ and $U_2(t)$, so that we can integrade the multiplier output for an appropriate interval of time without losing signal power.

4.1.2 Integration of the Multiplier Output

If we denote an instantaneous output of the multiplier as $y(t)$:

$$y(t) = U_1(t)U_2(t), \tag{310}$$

an output of the integrator (i.e., correlator output) $z(t)$ is related to the input $y(t)$ by a linear system of time averaging with an impulse response $a(t)$:

$$z(t) = y(t) * a(t) = \int_{-\infty}^{\infty} y(t')a(t-t')\,dt', \tag{311}$$

where symbol '$*$' stands for a convolution, as before.

If an input $y(t)$ is a constant y in time, the time averaging result must be equal to the constant y. Therefore, the impulse response $a(t)$ must satisfy:

$$\int_{-\infty}^{\infty} a(t)\, dt = 1. \tag{312}$$

For example, if we take a simple running mean with an averaging time $2T$:

$$z(t) = \frac{1}{2T} \int_{t-T}^{t+T} y(t')\, dt', \tag{313}$$

then the impulse response $a(t)$ is given by

$$a(t) = \begin{cases} \frac{1}{2T} & \text{for} \quad -T \le t \le T, \\ 0 & \text{otherwise.} \end{cases}$$

4.1.3 Statistical Expectation of the Correlator Output

We assume that the noise components from two antennas are not correlated with each other, and are not correlated with signal components, either. Then, we have

$$\langle N_1(t)\, N_2(t) \rangle = 0, \;\; \langle V_i(t)\, N_j(t) \rangle = 0 \;\; \text{for} \;\; i = 1,\, 2 \;\; \text{and} \;\; j = 1,\, 2, \tag{314}$$

where $\langle \;\; \rangle$ stands for the ensemble average, as before. From equation (314), we obtain

$$\langle y(t) \rangle = \langle U_1\, U_2 \rangle = \langle [V_1(t) + N_1(t)]\, [V_2(t) + N_2(t)] \rangle = \langle V_1(t)\, V_2(t) \rangle. \tag{315}$$

In next subsection, we will see that the multiplier output $y(t) = U_1(t)U_2(t)$ is a stationary random process, provided that $U_1(t)$ and $U_2(t)$ are jointly stationary random processes obeying the zero–mean jointly normal (or Gaussian) probability density. If so,

$$\langle y(t) \rangle = \text{const},$$

and, therefore,

$$\langle z(t) \rangle = \int_{-\infty}^{\infty} \langle y(t') \rangle a(t - t')\, dt' = \langle y(t) \rangle \int_{-\infty}^{\infty} a(t - t') dt' = \langle y(t) \rangle = \langle V_1(t)\, V_2(t) \rangle, \tag{316}$$

in view of equations (312) and (315). Equation (316) shows that the statistical expectation of the correlator output in a radio interferometer is equal to the expected correlation, i.e. the cross–correlation of signals $V_1(t)$ and $V_2(t)$, and does not contain system–noise contribution, unlike that of the square–law detector output of a single–dish radio telescope.

4.1.4 Dispersion of the Correlator Output

As we saw in the discussion of the correlation ergodicity in subsection 1.2.4, a correlator output $z(t)$ with a finite integration time should be associated with a noise–induced fluctuation around the expectation:

$$z(t) = \langle z(t) \rangle + \text{noise fluctuation} = \langle V_1(t)\, V_2(t) \rangle + \text{noise fluctuation}. \quad (317)$$

This fluctuation determines detection limit of a radio source as illustrated in Figure 82.

The magnitude of this fluctuation can be estimated by the dispersion σ_z^2 of the correlator output $z(t)$, as we saw in the discussion of the correlation ergotic processes. The dispersion is given by

$$\begin{aligned} \sigma_z^2 &= \langle [z(t) - \langle z(t) \rangle]^2 \rangle \\ &= \langle [z^2(t) - 2z(t)\, \langle z(t) \rangle + \langle z(t) \rangle^2] \rangle \\ &= \langle z^2(t) \rangle - \langle z(t) \rangle^2, \end{aligned} \quad (318)$$

and, therefore, the standard deviation is given by

$$\sigma_z = \sqrt{\langle z^2(t) \rangle - \langle z(t) \rangle^2}. \quad (319)$$

This is a measure of strength of the noise in the correlator output.

On the other hand, strength of the signal in the correlator output is given by the fringe amplitude $\mathcal{A}^U$ or $\mathcal{A}^L$ which are shown in equations (261) and (262) for USB and LSB receptions, respectively, for a continuum spectrum source.

Therefore, signal–to–noise ratio S/N for a continuum spectrum source is given by

$$S/N = \frac{\mathcal{A}}{\sigma_z}, \quad (320)$$

where $\mathcal{A}$ is either $\mathcal{A}^U$ or $\mathcal{A}^L$, depending on a choice of the single sideband actually received.

In following subsections, we will calculate the standard deviation of the correlator output σ_z, in order to get an explicit expression of the signal–to–noise ratio in terms of physical parameters characterizing the radio source and antenna–receiving systems.

4.2 Useful Formulae Related to the Correlator Output

In calculating the standard deviation of the correlator output σ_z, we will use following formulae relevant to the present problem.

4.2.1 Fourth Order Moment Relation

Let x_1, x_2, $\cdots$, x_n are zero–mean random variables obeying to the jointly normal (or Gaussian) probability density given in equation (30):

$$f(x_1, \cdots, x_n) = \frac{1}{\sqrt{(2\pi)^n \Delta}} e^{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i C_{ij}^{-1} x_j}, \tag{321}$$

where expectation η_i of the random variable x_i figuring in equation (30) is zero by assumption for all $i = 1, 2, \cdots, n$; $C_{ij} \equiv \langle x_i\, x_j \rangle$ is a covariance matrix, C_{ij}^{-1} is its inverse matrix, and $\Delta \equiv \det\{C_{ij}\}$ is its determinant. In general, we denote an inverse matrix by a symbol $(\ )^{-1}$ and a determinant by $\det(\ )$.

A very useful formula is known for such jointly normal random variables x_i $(i = 1, \cdots, n)$:

$$\langle x_i\, x_j\, x_k\, x_l \rangle = \langle x_i\, x_j \rangle \langle x_k\, x_l \rangle + \langle x_i\, x_k \rangle \langle x_j\, x_l \rangle + \langle x_i\, x_l \rangle \langle x_j\, x_k \rangle, \tag{322}$$

which says that a fourth order statistical moment (a term "n–th order statistical moment" is used here to denote a statistical expectation of a product of n random variables) $\langle x_i\, x_j\, x_k\, x_l \rangle$ is decomposed as a sum of products of second order statistical moments (correlations) $\langle x_i\, x_j \rangle$.

Proof :

1. Let us reformulate equation (321), using matrix notation of a row vector X and its transpose X^T:

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \text{and} \quad X^T = (x_1 \ \ x_2 \ \ \cdots \ \ x_n),$$

where $(\ )^T$ denotes a transpose matrix. Then we have

$$f(X) = \frac{1}{\sqrt{(2\pi)^n \Delta}} e^{-\frac{1}{2} X^T C^{-1} X}, \tag{323}$$

where C is the covariance matrix which is a positive–definite symmetric matrix with $C_{ij} = \langle x_i\, x_j \rangle$ as ij–element.

2. Fourier transform of a probability density is called "characteristic function". A characteristic function $F(\Omega)$ of the joint probability density $f(X)$ in equation (323), where Ω is a row vector of angular frequencies $\Omega^T = (\omega_1 \;\; \omega_2 \;\; \cdots \;\; \omega_n)$, is given by

$$\begin{aligned} F(\Omega) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(X)\, e^{-i\,\Omega^T X}\, dx_1 \cdots dx_n \;=\; \langle e^{-i\,\Omega^T X} \rangle \\ &= \frac{1}{\sqrt{(2\pi)^n \Delta}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2} X^T C^{-1} X - i\Omega^T X}\, dx_1 \cdots dx_n. \end{aligned} \tag{324}$$

3. According to the linear algebra, a symmetric matrix can be diagonalized by a suitable orthogonal transformation. So, let T be an orthogonal matrix, which diagonalizes the symmetric matrix C^{-1}:

$$T^T C^{-1} T = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}, \tag{325}$$

where $\lambda_1, \lambda_2, \cdots, \lambda_n$ are eigen values of the matrix C^{-1}.

Since T is an orthogonal matrix, by definition we have

$$T^T T = T T^T = I, \quad T^{-1} = T^T, \quad \det T = 1, \tag{326}$$

where I is a unit matrix:

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix},$$

and, therefore,

$$T^T C\, T = (T^T C^{-1} T)^{-1} = \begin{pmatrix} \lambda_1^{-1} & 0 & \cdots & 0 \\ 0 & \lambda_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^{-1} \end{pmatrix}, \tag{327}$$

and

$$\Delta = \det\, C = \det\, T^T\, \det\, C\, \det\, T = \det(T^T\, C\, T) = \frac{1}{\lambda_1\, \lambda_2\, \cdots\, \lambda_n}. \tag{328}$$

4. If we introduce a new row vector Y:

$$Y^T = (y_1\ \ y_2\ \ \cdots\ \ y_n),$$

which is obtained by an orthogonal transformation of X with T:

$$Y = T^T\, X,$$

then we have

$$X = T\, Y, \quad \frac{\partial x_i}{\partial y_j} = T_{ij}, \quad \frac{\partial(x_1\ x_2\ \cdots\ x_n)}{\partial(y_1\ y_2\ \cdots\ y_n)} = \det\, T = 1,$$

$$X^T\, C^{-1}\, X = Y^T\, T^T\, C^{-1}\, T\, Y = \sum_{j=1}^{n} \lambda_j\, y_j^2,$$

$$\Omega^T\, X = \Omega^T\, T\, Y = \sum_{i=1}^{n} \sum_{j=1}^{n} \omega_i\, T_{ij}\, y_j. \tag{329}$$

Therefore, the characteristic function $F(\Omega)$ given in equation (324) is reduced to

$$F(\Omega) = \frac{1}{\sqrt{(2\pi)^n \Delta}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{\ -\frac{1}{2} \sum_{j=1}^{n} \lambda_j\, y_j^2\ -\ i \sum_{i=1}^{n} \sum_{j=1}^{n} \omega_i\, T_{ij}\, y_j}\ dy_1 \cdots dy_n. \tag{330}$$

If we further introduce new variables:

$$z_j = \sqrt{\frac{\lambda_j}{2}}\, y_j \qquad (j = 1, \cdots, n),$$

then we have

$$\begin{aligned} F(\Omega) &= \frac{1}{\sqrt{(2\pi)^n \Delta}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \\ &\quad e^{\ -\sum_{j=1}^{n} (z_j^2 + i \sum_{i=1}^{n} \omega_i\, T_{ij} \sqrt{\frac{2}{\lambda_j}}\, z_j)}\ \sqrt{\frac{2^n}{\lambda_1 \cdots \lambda_n}}\, dz_1 \cdots dz_n, \\ &= \frac{1}{\sqrt{\pi^n}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{\ -\sum_{j=1}^{n} (z_j^2 + i \sum_{i=1}^{n} \omega_i\, T_{ij} \sqrt{\frac{2}{\lambda_j}}\, z_j)}\ dz_1 \cdots dz_n. \end{aligned} \tag{331}$$

5. Applying the integration formula:

$$\int_{-\infty}^{\infty} e^{-x^2 - i\,a\,x}\,dx = \sqrt{\pi}\,e^{-\frac{a^2}{4}},$$

we reduce equation (331) to

$$\begin{aligned} F(\Omega) &= e^{-\frac{1}{4}\sum_{j=1}^{n}\left(\sum_{i=1}^{n}\omega_i\,T_{ij}\sqrt{\frac{2}{\lambda_j}}\right)^2} \\ &= e^{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\omega_i\,T_{ij}\,\frac{1}{\lambda_j}\,T_{kj}\,\omega_k} \\ &= e^{-\frac{1}{2}\,\Omega^T\,C\,\Omega}. \end{aligned} \tag{332}$$

6. From equations (324) and (332), we have

$$\langle e^{-i\,\Omega^T\,X}\rangle = e^{-\frac{1}{2}\,\Omega^T\,C\,\Omega}. \tag{333}$$

Expanding exponential functions in both sides of this equation into the Taylor series, we have

$$\begin{aligned} \langle e^{-i\Omega^T\,X}\rangle &= 1 - \frac{1}{2!}\,\langle(\Omega^T\,X)^2\rangle + i\frac{1}{3!}\,\langle(\Omega^T\,X)^3\rangle + \frac{1}{4!}\,\langle(\Omega^T\,X)^4\rangle - \cdots, \\ e^{-\frac{1}{2}\,\Omega^T\,C\,\Omega} &= 1 - \frac{1}{2}\,\Omega^T\,C\,\Omega + \frac{1}{2!}\,(\frac{1}{2}\,\Omega^T\,C\,\Omega)^2 - \cdots. \end{aligned} \tag{334}$$

Terms with fourth order of ω_i's in the above equations are

$$\frac{1}{4!}\,\langle(\Omega^T\,X)^4\rangle = \frac{1}{4!}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\langle x_i\,x_j\,x_k\,x_l\rangle\omega_i\,\omega_j\,\omega_k\,\omega_l, \tag{335}$$

and

$$\begin{aligned} \frac{1}{2!}\,(\frac{1}{2}\,\Omega^T\,C\,\Omega)^2 &= \frac{1}{2!}\,\frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}C_{ij}\,C_{kl}\,\omega_i\,\omega_j\,\omega_k\,\omega_l \\ &= \frac{1}{4!}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}(C_{ij}\,C_{kl} + C_{ik}\,C_{jl} + C_{il}\,C_{jk})\,\omega_i\,\omega_j\,\omega_k\,\omega_l, \end{aligned} \tag{336}$$

respectively. Right hand sides of equations (335) and (336) are equal to each other in view of equation (333). Consequently, coefficients of $\omega_i\,\omega_j\,\omega_k\,\omega_l$ in these equations, i.e., $\langle x_i\,x_j\,x_k\,x_l\rangle$ and $C_{ij}\,C_{kl} + C_{ik}\,C_{jl} + C_{il}\,C_{jk}$, must be equal to each other, too, since both of them are invariant in any substitution of i, j, k, l. Thus, we have

$$\langle x_i\,x_j\,x_k\,x_l\rangle = C_{ij}\,C_{kl} + C_{ik}\,C_{jl} + C_{il}\,C_{jk}. \tag{337}$$

Since $C_{ij} = \langle x_i\,x_j\rangle$, this completes the proof of equation (322).

4.2.2 Multiplier Output as a Stationary Random Process

If two voltage inputs $U_1(t)$ and $U_2(t)$ of a correlator are jointly stationary processes and if they obey the zero–mean jointly normal (or Gaussian) probability density such as given in equation (321), then their product $y(t) = U_1(t)\,U_2(t)$ is also a stationary random process.

Proof :

1. Expectation of $y(t)$ is a constant in time. In fact,

$$\langle y(t)\rangle = \langle U_1(t)U_2(t)\rangle = R_{U_1U_2}(0), \tag{338}$$

   where $R_{U_1U_2}(\tau) = \langle U_1(t)\,U_2(t-\tau)\rangle$ is a cross–correlation of jointly stationary random processes $U_1(t)$ and $U_2(t)$, and, therefore, $\langle y(t)\rangle$ does not depend on time t.

2. Autocorrelation of $y(t)$ is a function of time difference only. In fact, in view of equation (322), we have

$$\begin{aligned}\langle y(t)\,y(t-\tau)\rangle &= \langle U_1(t)\,U_2(t)\,U_1(t-\tau)\,U_2(t-\tau)\rangle \\ &= \langle U_1(t)\,U_2(t)\rangle\,\langle U_1(t-\tau)\,U_2(t-\tau)\rangle \\ &\quad +\langle U_1(t)\,U_1(t-\tau)\rangle\,\langle U_2(t)\,U_2(t-\tau)\rangle \\ &\quad +\langle U_1(t)\,U_2(t-\tau)\rangle\,\langle U_2(t)\,U_1(t-\tau)\rangle \\ &= R^2_{U_1U_2}(0) + R_{U_1U_1}(\tau)\,R_{U_2U_2}(\tau) + R_{U_1U_2}(\tau)\,R_{U_1U_2}(-\tau),\end{aligned} \tag{339}$$

   where $R_{U_1U_1}(\tau) = \langle U_1(t)\,U_1(t-\tau)\rangle$ and $R_{U_2U_2}(\tau) = \langle U_2(t)\,U_2(t-\tau)\rangle$ are autocorrelations of input voltages $U_1(t)$ and $U_2(t)$. This equation shows that $\langle y(t)\,y(t-\tau)\rangle$ is a function of time difference τ only.

Thus, we proved that the multiplier output $y(t)$ is a stationary random process.

Since the integrator output (i.e., correlator output) $z(t)$ is an output of a linear system of time averaging, given in equation (311), with the input $y(t)$ which was shown to be a stationary random process, we conclude that $z(t)$ is also a stationary random process.

Hereafter, we will denote autocorrelations of the multiplier output $y(t)$ and the integrator output $z(t)$ as functions of time difference τ:

$$\begin{aligned}R_{yy}(\tau) &= \langle y(t)\,y(t-\tau)\rangle, \\ R_{zz}(\tau) &= \langle z(t)\,z(t-\tau)\rangle.\end{aligned} \tag{340}$$

4.2.3 Time Averaging Operator

Since the multiplier output $y(t)$ and the integrator output $z(t)$ are related to each other through the impulse response $a(t)$ of the operator of time averaging, as shown in equation (311):

$$z(t) = y(t) * a(t),$$

their autocorrelations are related to each other as:

$$R_{zz}(\tau) = R_{yy}(\tau) * a(\tau) * a(-\tau), \tag{341}$$

in view of equation (43).

Therefore, if we denote power spectra of $y(t)$ and $z(t)$ as $S_{yy}(\omega)$ and $S_{zz}(\omega)$:

$$S_{yy}(\omega) \Leftrightarrow R_{yy}(\tau),$$
$$S_{zz}(\omega) \Leftrightarrow R_{zz}(\tau),$$

where symbol $\Leftrightarrow$ stands for a Fourier transformation pair, and introduce a system function $\mathsf{A}(\omega)$ of $a(t)$:

$$\mathsf{A}(\omega) \Leftrightarrow a(t),$$

then we have, in view of equation (75),

$$S_{zz}(\omega) = S_{yy}(\omega) \mid \mathsf{A}(\omega) \mid^2 . \tag{342}$$

Since

$$\mathsf{A}(\omega) = \int_{-\infty}^{\infty} a(t)\, e^{-i\omega t} dt,$$

and, from equation (312),

$$\int_{-\infty}^{\infty} a(t)\, dt = 1,$$

the system function $\mathsf{A}(\omega)$ must always satisfy an equation:

$$\mathsf{A}(0) = 1. \tag{343}$$

Note that, for the case of a simple running mean, the system function takes the sinc function form:

$$\mathsf{A}(\omega) = \frac{1}{2T} \int_{-T}^{T} e^{-i\omega t} dt = \frac{\sin(\omega T)}{\omega\, T}. \tag{344}$$

In general, $\mathsf{A}(\omega)$ can be regarded as a low–pass filter with a narrow passband of about $\pm 2\pi/T$, as evident from the above simple case. If the integral time T is 1 s, for example, the passband is as narrow as ± 1 Hz.

4.2.4 Power Spectrum of Multiplier Output

Let us now describe the power spectrum of the multiplier output $S_{yy}(\omega)$ through power and cross–power spectra of voltages, which are the inputs to our correlator.

From equations (339) and (340), we have

$$R_{yy}(\tau) = R^2_{U_1U_2}(0) + R_{U_1U_1}(\tau)\, R_{U_2U_2}(\tau) + R_{U_1U_2}(\tau)\, R_{U_1U_2}(-\tau). \qquad (345)$$

Therefore, Fourier transformation of this equation gives us the power spectrum $S_{yy}(\omega)$. Using convolution theorem in Fourier transformation, given in equation (68), for products of functions of τ, we obtain

$$\begin{aligned} S_{yy}(\omega) &= \int_{-\infty}^{\infty} R_{yy}(\tau) e^{-i\omega\tau} d\tau \\ &= 2\pi R^2_{U_1U_2}(0)\, \delta(\omega) \\ &\quad + \frac{1}{2\pi} S_{U_1U_1}(\omega) * S_{U_2U_2}(\omega) + \frac{1}{2\pi} S_{U_1U_2}(\omega) * S^*_{U_1U_2}(\omega), \end{aligned} \qquad (346)$$

where $\delta(\omega)$ is the delta function of angular frequency ω, and $S_{U_1U_1}(\omega)$, $S_{U_2U_2}(\omega)$, and $S_{U_1U_2}(\omega)$ are power and cross–power spectra of input voltages:

$$\begin{aligned} S_{U_1U_1}(\omega) &\Leftrightarrow R_{U_1U_1}(\tau), \\ S_{U_2U_2}(\omega) &\Leftrightarrow R_{U_2U_2}(\tau), \\ S_{U_1U_2}(\omega) &\Leftrightarrow R_{U_1U_2}(\tau). \end{aligned}$$

Here we used a relation:

$$\int_{-\infty}^{\infty} e^{-i\omega\tau} d\tau = 2\pi\delta(\omega),$$

which was given in equation (66), and for Fourier transform of a real function $R_{U_1U_2}(-\tau)$:

$$\begin{aligned} \int_{-\infty}^{\infty} R_{U_1U_2}(-\tau) e^{-i\omega\tau} d\tau &= \int_{-\infty}^{\infty} R_{U_1U_2}(\tau) e^{i\omega\tau} d\tau \\ &= \left(\int_{-\infty}^{\infty} R_{U_1U_2}(\tau) e^{-i\omega\tau} d\tau \right)^* = S^*_{U_1U_2}(\omega). \end{aligned}$$

4.2.5 Power and Cross–Power Spectra of Input Voltages

Power and cross–power spectra of input voltages figuring in equation (346) can be described as follows.

Assuming, as before, that signal and noise are not correlated with each other, and noises from different antennas are not correlated, either, we express autocorrelations and cross–correlation of input voltages $U_1(t)$ and $U_2(t)$ in terms of signal (V) and noise (N) components:

$$\begin{aligned} R_{U_1U_1}(\tau) &= R_{V_1V_1}(\tau) + R_{N_1N_1}(\tau), \\ R_{U_2U_2}(\tau) &= R_{V_2V_2}(\tau) + R_{N_2N_2}(\tau), \\ R_{U_1U_2}(\tau) &= R_{V_1V_2}(\tau), \end{aligned} \tag{347}$$

where

$$\begin{aligned} R_{V_1V_1}(\tau) &= \langle V_1(t)\, V_1(t-\tau)\rangle, \\ R_{V_2V_2}(\tau) &= \langle V_2(t)\, V_2(t-\tau)\rangle, \\ R_{V_1V_2}(\tau) &= \langle V_1(t)\, V_2(t-\tau)\rangle, \end{aligned}$$

are auto– and cross–correlations of signal voltages $V_1(t)$ and $V_2(t)$, while

$$\begin{aligned} R_{N_1N_1}(\tau) &= \langle N_1(t)\, N_1(t-\tau)\rangle, \\ R_{N_2N_2}(\tau) &= \langle N_2(t)\, N_2(t-\tau)\rangle, \end{aligned}$$

are autocorrelations of noise voltages $N_1(t)$ and $N_2(t)$.

Therefore, the power and cross–power spectra are obtained by Fourier transformation of equations (347) as

$$\begin{aligned} S_{U_1U_1}(\omega) &= S_{V_1V_1}(\omega) + S_{N_1N_1}(\omega), \\ S_{U_2U_2}(\omega) &= S_{V_2V_2}(\omega) + S_{N_2N_2}(\omega), \\ S_{U_1U_2}(\omega) &= S_{V_1V_2}(\omega), \end{aligned} \tag{348}$$

where

$$\begin{aligned} S_{V_1V_1}(\omega) &\Leftrightarrow R_{V_1V_1}(\tau), \\ S_{V_2V_2}(\omega) &\Leftrightarrow R_{V_2V_2}(\tau), \\ S_{V_1V_2}(\omega) &\Leftrightarrow R_{V_1V_2}(\tau), \end{aligned}$$

are power and cross–power spectra of signal voltages, while

$$\begin{aligned} S_{N_1N_1}(\omega) &\Leftrightarrow R_{N_1N_1}(\tau), \\ S_{N_2N_2}(\omega) &\Leftrightarrow R_{N_2N_2}(\tau), \end{aligned}$$

are power spectra of noise voltages.

4.2.6 Antenna Temperature and System Noise Temperature

We saw in equation (169), that a power spectrum $S_{v_iv_i}(\omega)$ of a received signal voltage $v_i(t)$ $(i = 1, 2)$ at RF band, which is just received by i–th antenna, and has not yet gone through a receiving system, is given in terms of effective flux density $\mathcal{S}_\nu$ of an observed source and efective aperture of i–th antenna A_{e_i} by

$$S_{v_iv_i}(\omega) = \frac{1}{4} A_{e_i} \mathcal{S}_\nu \quad (i = 1, 2).$$

Since the effective flux density and the antenna temperature $T_{A_i}(\omega)$, characterizing a received power from a radio source per unit bandwidth, are related to each other by an equation:

$$k\, T_{A_i}(\omega) = \frac{1}{2} A_{e_i} \mathcal{S}_\nu,$$

as we saw in Chapter 2, where $k = 1.381 \times 10^{-23}$ JK^{-1} is Boltzmann's constant, the power spectrum of the received signal voltage is described through the antenna temperature as:

$$S_{v_iv_i}(\omega) = \frac{1}{2}\, k\, T_{A_i}(\omega) \quad (i = 1, 2). \tag{349}$$

Since the frequency conversion preserves the power spectrum of the received signal voltage, we can easily derive from equation (349) a power spectrum $S_{V_iV_i}(\omega)$ of an IF signal voltage $V_i(t)$ at an input of a correlator. Indeed, if we denote the system function of the receiving system as $H_i(\omega)$, as given in equation (195), and if we assume for definiteness upper sideband (USB) reception, we have

$$S_{V_iV_i}(\omega) = \frac{1}{2}\, k\, T_{A_i}(\omega_{LO} + \omega) \mid H_i(\omega) \mid^2 \quad (\text{for } \omega \geq 0,\ i = 1, 2), \tag{350}$$

where ω is frequency in IF band and ω_{LO} is local oscillator frequency of the frquency conversion. Here we showed only positive frequency side of the power spectrum, which is an essentially even function of frequency.

On the other hand, the noise component is described through the input equivalent system noise temperature $T_{S_i}(\omega)$, which is supposed to pass through the same receiving system as the radio source signal does. Thus, in the USB reception and in the positive frequency range, we have

$$S_{N_iN_i}(\omega) = \frac{1}{2}\, k\, T_{S_i}(\omega_{LO} + \omega) \mid H_i(\omega) \mid^2 \quad (\text{for } \omega \geq 0,\ i = 1, 2). \tag{351}$$

Therefore, from equations (348), (350), and (351), the power spectrum of the input voltage $U_i(t)$ is given by

$$S_{U_iU_i}(\omega) = \frac{1}{2}\, k\, [T_{A_i}(\omega_{LO}+\omega)+T_{S_i}(\omega_{LO}+\omega)]\ \mid H_i(\omega)\mid^2 \quad (\text{for } \omega \geq 0,\ i = 1, 2), \tag{352}$$

in the USB reception and in the positive frequency range.

The cross–power spectrum of two input signal voltages $V_i(t),\ (i = 1, 2)$ after delay tracking and fringe stopping is given in equation (242), for the case of the USB reception and in the positive frequency range of the Hermitian symmetric cross–power spectrum. Therefore, taking into account also equation (348), the cross–power spectrum of the input voltages $U_i(t),\ (i = 1, 2)$ in the positive frequency range ($\omega \geq 0$) is given by

$$S_{U_1U_2}(\omega) = S_{V_1V_2}(\omega) = \frac{1}{4}\, A_0\, e^{-i(\phi_{LO1}-\phi_{LO2})}\, \mathcal{V}(\omega_{LO}+\omega)\, H_1(\omega)\, H_2^*(\omega), \tag{353}$$

where A_0 is geometric mean of effective apertures of antennas, $\phi_{LO1}-\phi_{LO2}$ is difference of initial phases of local oscillators, and we put $\Delta\tau_g = 0$ in equation (242), assuming complete delay tracking and fringe stopping.

In a simple case of a point–like radio source with normalized power pattern of the interferometer $A_N(\boldsymbol{\sigma}) = 1$ at centers of antenna beams, the complex visibility of the source is real and equal to its flux density, $\mathcal{V}(\omega_{LO}+\omega) = \mathcal{S}_\nu$, as given in equation (295), and, therefore,

$$\frac{1}{4}\, A_0\, \mathcal{V}(\omega_{LO}+\omega) = \frac{1}{4}\, \sqrt{A_{e_1}A_{e_2}}\, \mathcal{S}_\nu = \frac{1}{2}\, k\, \sqrt{T_{A_1}(\omega_{LO}+\omega)\, T_{A_2}(\omega_{LO}+\omega)}. \tag{354}$$

In this case, we have, in $\omega \geq 0$

$$S_{U_1U_2}(\omega) = \frac{1}{2}\, k\, \sqrt{T_{A_1}(\omega_{LO}+\omega)\, T_{A_2}(\omega_{LO}+\omega)}\ e^{-i(\phi_{LO1}-\phi_{LO2})}\, H_1(\omega)\, H_2^*(\omega). \tag{355}$$

4.3 Sensitibity of a Radio Interferometer

4.3.1 Standard Deviation Due to the Noise

Now we are ready to calculate the standard deviation σ_z of the correlator output $z(t)$ in equation (319), and then the signal–to–noise ratio in equation (320).

In equation (319):

$$\sigma_z = \sqrt{\langle z^2(t)\rangle - \langle z(t)\rangle^2},$$

the first term $\langle z^2(t)\rangle$ in square root is given by inverse Fourier transformation at $\tau = 0$ of the power spectrum $S_{zz}(\omega)$, which is related to $S_{yy}(\omega)$ given in equation (346) through equation (342). Therefore, we have

$$\begin{aligned}
\langle z^2(t)\rangle &= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{zz}(\omega)\, d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{yy}(\omega)\ \mid \mathsf{A}(\omega)\mid^2\ d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\Big[2\pi R^2_{U_1U_2}(0)\,\delta(\omega) \\
&+ \frac{1}{2\pi}S_{U_1U_1}(\omega) * S_{U_2U_2}(\omega) + \frac{1}{2\pi}S_{U_1U_2}(\omega) * S^*_{U_1U_2}(\omega)\Big]\ \mid \mathsf{A}(\omega)\mid^2\ d\omega \\
&= R^2_{U_1U_2}(0) \\
&+\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\Big[S_{U_1U_1}(\omega) * S_{U_2U_2}(\omega) + S_{U_1U_2}(\omega) * S^*_{U_1U_2}(\omega)\Big]\ \mid \mathsf{A}(\omega)\mid^2\ d\omega,
\end{aligned} \tag{356}$$

where $\mathsf{A}(\omega)$ is the system function of time averaging, and we used equation (343) for deriving the first term. On the other hand, second term $\langle z(t)\rangle^2$ is equal to $R^2_{U_1U_2}(0)$, since, in view of equations (317) and (338), we have

$$\langle z(t)\rangle = R_{U_1U_2}(0).$$

Therefore, the first term of the right hand side of equation (356) is compensated by $\langle z(t)\rangle^2$, and hence we have

$$\sigma_z^2 = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\Big[S_{U_1U_1}(\omega) * S_{U_2U_2}(\omega) + S_{U_1U_2}(\omega) * S^*_{U_1U_2}(\omega)\Big]\ \mid \mathsf{A}(\omega)\mid^2\ d\omega. \tag{357}$$

If we use explicit forms of convolution integrals, equation (357) is reduced to

$$\begin{aligned}
\sigma_z^2 = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}&[S_{U_1U_1}(\omega-\omega')\, S_{U_2U_2}(\omega') \\
&+S_{U_1U_2}(\omega-\omega')\, S^*_{U_1U_2}(\omega')\Big]\ \mid \mathsf{A}(\omega)\mid^2\ d\omega'\, d\omega.
\end{aligned} \tag{358}$$

Since $\mathsf{A}(\omega)$ is a very narrow–band low–pass filter around $\omega = 0$, as we saw earlier, $S_{U_1U_1}(\omega-\omega')$ and $S_{U_1U_2}(\omega-\omega')$ can be replaced by $S_{U_1U_1}(-\omega') = S_{U_1U_1}(\omega')$ and $S_{U_1U_2}(-\omega') = S^*_{U_1U_2}(\omega')$, and we can take them out of the integration with respect to ω. Thus we obtain,

$$\sigma_z^2 = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\Big[S_{U_1U_1}(\omega')\, S_{U_2U_2}(\omega') + S^*_{U_1U_2}(\omega')\, S^*_{U_1U_2}(\omega')\Big]\ d\omega'\int_{-\infty}^{\infty}\ \mid \mathsf{A}(\omega)\mid^2\ d\omega. \tag{359}$$

Taking into account that $S_{U_1U_1}(\omega')$ is an even, and $S_{U_1U_2}(\omega')$ is a Hermitian symmetric, functions, we can further reduce equation (359) to

$$\sigma_z^2 = \frac{1}{2\pi^2} \int_0^\infty \{S_{U_1U_1}(\omega')\, S_{U_2U_2}(\omega') + \Re\left[S_{U_1U_2}(\omega')\, S_{U_1U_2}(\omega')\right]\}\, d\omega' \int_{-\infty}^{\infty} \mid \mathsf{A}(\omega) \mid^2\, d\omega. \tag{360}$$

Now, assuming the USB reception, for definiteness, we apply equations (352) and (355) for spectra in positive frequency range ($\omega \geq 0$) of input voltages to obtain

$$\sigma_z^2 = \frac{k^2}{8\pi^2} \int_0^\infty \Big\{(T_{A_1} + T_{S_1})\,(T_{A_2} + T_{S_2})\, \mid H_1(\omega') \mid^2 \mid H_2(\omega') \mid^2 + \Re\left[(T_{A_1}\, T_{A_2})\, e^{-2i(\phi_{LO_1} - \phi_{LO_2})}\, H_1(\omega')\, H_2^*(\omega')\, H_1(\omega')\, H_2^*(\omega')\right]\Big\}\, d\omega' \times \int_{-\infty}^{\infty} \mid \mathsf{A}(\omega) \mid^2\, d\omega, \tag{361}$$

where we omitted argument $\omega_{LO} + \omega$ in T_{A_i} and T_{S_i} $(i = 1, 2)$ for simplicity.

Since, in actual observations of most of radio sources, antenna temperatures are much smaller than system noise temperatures: $T_{A_i} \ll T_{S_i}$ $(i = 1, 2)$, we ignore terms with T_{A_i}, compared with terms with T_{S_i} $(i = 1, 2)$. Then, assuming a flat noise spectrum, we obtain an equation:

$$\sigma_z^2 = \frac{k^2\, T_{S_1}\, T_{S_2}}{8\pi^2} \int_0^\infty \mid H_1(\omega') \mid^2 \mid H_2(\omega') \mid^2\, d\omega' \int_{-\infty}^{\infty} \mid \mathsf{A}(\omega) \mid^2\, d\omega, \tag{362}$$

which describes the dispersion of noise fluctuations in the correlator output.

4.3.2 Signal in Correlator Output

Signal part of the correlator output is equal to the cross–correlation $R_{U_1U_2}(\tau)$ of input voltages at $\tau = 0$ after delay tracking and fringe stopping, which is derived by inverse Fourier transformation of cross–power spectrum of the voltages with $\tau = 0$, as we saw in Subsection 2.9.1. So, if we assume a continuum spectrum source with constant visibility $\mathcal{V}$ with respect to ω, the signal in the correlator output, that is the cross–correlation $R_{U_1U_2}(0)$, is obtained from the cross–power spectrum $S_{U_1U_2}(\omega)$ given in equation (353) by

inverse Fourier transformation with zero argument:

$$R_{U_1U_2}(0) = \frac{1}{\pi}\,\Re \int_0^\infty S_{U_1U_2}(\omega)\,d\omega = A_0\,\Re[e^{-i(\phi_{LO1}-\phi_{LO2})}\,\mathcal{V}\,\mathcal{B}_{12}],$$

where A_0 is the geometric mean of effective apertures of element antennas, and

$$\mathcal{B}_{12} = \frac{1}{4\pi}\int_0^\infty H_1(\omega)H_2^*(\omega)\,d\omega,$$

is the bandwidth pattern as defined in equation (254) with $\Delta\tau_g = 0$ (perfect delay tracking and fringe stopping). Therefore, amplitude $\mathcal{A}$ of the signal part $R_{U_1U_2}(0)$ in the correlator output, is given by

$$\mathcal{A} = A_0 \mid \mathcal{V} \mid \mid \mathcal{B}_{12} \mid .$$

Note that this is just the theoretical expression of the fringe amplitude given in equation (261), as expected.

In a simple case of a point–like source, when the visibility $\mathcal{V}$ is equal to the flux density $\mathcal{S}_\nu$, we have $A_0 \mid \mathcal{V} \mid = \sqrt{A_{e_1}\,A_{e_2}}\,\mathcal{S}_\nu = 2\,k\,\sqrt{T_{A_1}\,T_{A_2}}$, as we saw in equation (354). Therefore, we obtain

$$\mathcal{A} = 2\,k\,\sqrt{T_{A_1}\,T_{A_2}}\ \mid \mathcal{B}_{12} \mid, \tag{363}$$

for the fringe amplitude, or the amplitude of the signal part $R_{U_1U_2}(0$ in the correlator output.

4.3.3 Signal–to–Noise Ratio of the Correlator Output

Consequently, the signal–to–noise ratio is given by

$$\begin{aligned} S/N &= \frac{\mathcal{A}}{\sigma_z} \\ &= 8\pi\sqrt{\frac{T_{A_1}\,T_{A_2}}{T_{S_1}\,T_{S_2}}}\;\frac{\mid \mathcal{B}_{12} \mid}{\sqrt{2\int_0^\infty \mid H_1(\omega') \mid^2 \mid H_2(\omega') \mid^2\ d\omega'}}\;\frac{1}{\sqrt{\int_{-\infty}^{\infty} \mid \mathsf{A}(\omega) \mid^2\ d\omega}}. \end{aligned} \tag{364}$$

Let us first assume that $H_i(\omega)$ (i =1 or 2) is a real and positive filter of simple rectangular form given by

$$\mid H_i(\omega) \mid^2 = H_i^2(\omega) = \begin{cases} G_i & \text{if } \ \omega_I - \frac{\Delta\omega}{2} \le \omega \le \omega_I + \frac{\Delta\omega}{2}, \\ 0 & \text{otherwise,} \end{cases} \tag{365}$$

in the positive frequency range ($\omega \geq 0$), where G_i is a gain factor in a receiving system of i—th antenna, ω_I is a center IF frequency of the filter passband, and $\Delta\omega = 2\pi B$ is an angular frequency bandwidth, corresponding to a frequency bandwidth B.

In this case, it is evident from equation (254) that the coefficient G of the bandwidth pattern as given in equation (260) is equal to $G = \sqrt{G_1\, G_2}$. Therefore, the amplitude of the bandwidth pattern $\mid \mathcal{B}_{12} \mid$ is reduced to

$$\mathcal{B}_{12} = \frac{\sqrt{G_1\, G_2}B}{2}\, \frac{\sin(\pi B \Delta\tau_g)}{\pi B \Delta\tau_g} = \frac{\sqrt{G_1\, G_2}B}{2} \quad (\text{for } \Delta\tau_g = 0), \tag{366}$$

where we assumed complete delay tracking and fringe stopping with residual delay $\Delta\tau_g = 0$. On the other hand, we obviously have

$$\int_0^\infty \mid H_1(\omega') \mid^2 \mid H_2(\omega') \mid^2 \, d\omega' = 2\pi\, G_1\, G_2\, B. \tag{367}$$

Therefore, we obtain

$$\frac{8\pi \mid \mathcal{B}_{12} \mid}{\sqrt{2 \int_0^\infty \mid H_1(\omega') \mid^2 \mid H_2(\omega') \mid^2 \, d\omega'}} = \sqrt{4\pi\, B}. \tag{368}$$

For more general case of non–rectangular filters, we will adopt this equation (368) as a definition of bandwidth B.

On the other hand, if we assume the simple running mean of equation (313) for the time averaging, then an integration time τ_a is equal to the interval of averaging $\tau_a = 2\,T$. In this case, equation (344) shows that the system function of time averaging $\mathsf{A}(\omega)$ has a simple sinc function form:

$$\mathsf{A}(\omega) = \frac{\sin(\omega\, T)}{\omega\, T}.$$

Therefore,

$$\int_{-\infty}^\infty \mid \mathsf{A}(\omega) \mid^2 \, d\omega = \int_{-\infty}^\infty \frac{\sin^2(\omega\, T)}{\omega^2\, T^2}\, d\omega = \frac{\pi}{T} = \frac{2\pi}{\tau_a},$$

and, hence, we have

$$\frac{1}{\sqrt{\int_{-\infty}^\infty \mid \mathsf{A}(\omega) \mid^2 \, d\omega}} = \sqrt{\frac{\tau_a}{2\pi}}. \tag{369}$$

For more general case of non–running–mean averaging, we will adopt this equation (369) as a definition of integration time τ_a.

Inserting equations (368) and (369) to equation (364), we obtain an equation for the signal–to–noise ratio S/N in radio interferometry:

$$S/N = \frac{\mathcal{A}}{\sigma_z} = \sqrt{\frac{T_{A_1}\,T_{A_2}}{T_{S_1}\,T_{S_2}}}\,\sqrt{2\,B\,\tau_a}. \tag{370}$$

It is worthwhile to note that, in our approximation of a white spectrum point–like source and simple rectangular filters, the ratio of geometric means of antenna temperatures and that of system noise temperatures in equation (370) is equal to maximum amplitude of cross–correlation coefficient $r_{U_1U_2}(\tau)$ of input voltages U_1 and U_2 at $\tau = 0$:

$$r_{U_1U_2}(0) = \frac{R_{U_1U_2}(0)}{\sqrt{R_{U_1U_1}(0)\,R_{U_2U_2}(0)}},$$

achieved when delay tracking and fringe stopping are perfectly performed.

In fact, in view of equations (363) and (366), amplitude $\mathcal{A}$ of $R_{U_1U_2}(0)$, that is the fringe amplitude, is given by

$$\mathcal{A} = k\sqrt{T_{A_1}\,T_{A_2}}\,\sqrt{G_1\,G_2}B.$$

On the other hand, the autocorrelation at zero argument, or the power, $R_{U_iU_i}(0)$, with $i = 1,\ 2$, is derived from power spectrum $S_{U_iU_i}(\omega) \cong S_{N_iN_i}(\omega)$ (since $T_{A_i} \ll T_{S_i}$) in equations (352) and (351):

$$R_{U_iU_i}(0) \cong \frac{1}{\pi}\int_0^\infty S_{N_iN_i}(\omega)\,d\omega = \frac{1}{2\pi}\,k\,T_{S_i}\int_0^\infty |H_i(\omega)|^2\,d\omega = k\,T_{S_i}\,G_i\,B.$$

Therefore, the maximum amplitude of the cross–correlation coefficeint $r_{U_1U_2}(0)$ is really given by

$$\frac{\mathcal{A}}{\sqrt{R_{U_1U_1}(0)\,R_{U_2U_2}(0)}} = \sqrt{\frac{T_{A_1}\,T_{A_2}}{T_{S_1}\,T_{S_2}}}.$$

4.3.4 Additional Remarks on the S/N Formula

Although we derived equation (370) assuming a point–like and continuum spectrum source in a case of USB reception, the equation can be applied to more general cases, if we

- multiply a ratio of correlated flux density $\mid \mathcal{V} \mid$ and total flux density $\mathcal{S}_{T_\nu}$, i.e., $\mid \mathcal{V} \mid / \mathcal{S}_{T_\nu}$, to $\sqrt{T_{A_1}\, T_{A_2}}$ for an extended source,

- no change is required in equation (370) for a spectralline source as far as T_{A_1} and T_{A_2} are average antenna temperatures over a whole filter bandwidth B, or multiply a ratio of line width b and B, i.e., b/B, if T_{A_1} and T_{A_2} are peak antenna temperatures confined within the line,

- take LSB values of T_{A_i} and T_{S_i} $(i = 1, 2)$ in LSB reception case.

An extra efficiency factor η_c called "coherence factor", typically $\eta_c \cong 0.5 \sim 0.9$, is multiplied to the right hand side of equation (370) in case of digital data processing and VLBI. This factor takes into account losses associated with digitization, digital logics, fringe–phase fluctuations due to instability of frequency standards and/or atmosphere, and imperfection of theoretically predicted delay model, as we will discuss in Chapter 4. Then, the signal–to–noise ratio S/N for VLBI is usually given by:

$$S/N = \eta_c \sqrt{\frac{T_{A_1}\, T_{A_2}}{T_{S_1}\, T_{S_2}}}\, \sqrt{2\, B\, \tau_a}. \tag{371}$$

4.3.5 A Simple Interpretation of the $\sqrt{B\,\tau_a}$ Factor

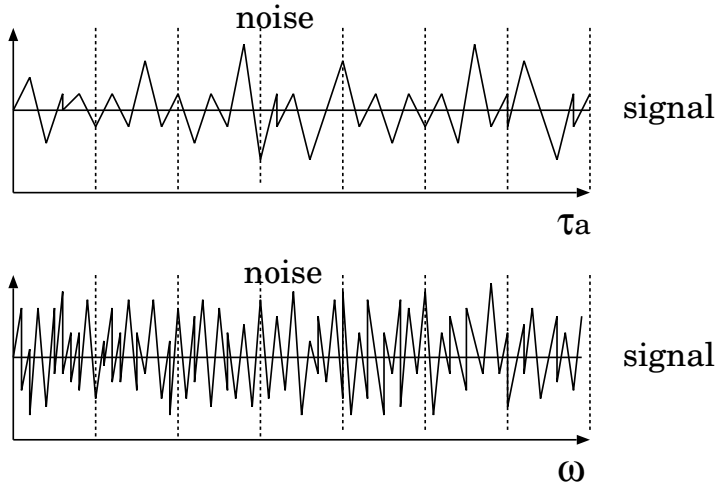

Figure 84: Schematic diagram explaining how the signal–to–noise ratio of a stationary signal with a continuum spectrum is improved by increasing the integration time τ_a and the bandwidth B, in proportion to $\sqrt{B\,\tau_a}$.

Figure 84 gives a qualitative explanation why we have a factor $\sqrt{B\,\tau_a}$ in a formula of signal–to–noise ratio. In fact, we can increase the number of independent measurements, by increasing the integration time τ_a and the bandwidth B for a stationary source with a continuum spectrum, thus improving the signal–to–noise ratio in proportion to $\sqrt{\tau_a}$ and $\sqrt{B}$.

4.3.6 Formulae for S/N and Minimum Detectable Flux Density

If correlated flux density of an observed source is $\mathcal{S}_\nu$, and effective apertures, aperture diameters, and aperture efficiencies of antennas are A_{e1}, A_{e2}, L_1, L_2, and η_{A_1}, η_{A_2}, respectively, we can rewrite equation (371) for the signal–to–noise ratio S/N in a new form:

$$S/N = \eta_c \frac{\sqrt{A_{e_1}\,A_{e_2}}\,\mathcal{S}_\nu}{2k\,\sqrt{T_{S_1}\,T_{S_2}}}\,\sqrt{2\,B\,\tau_a}, \tag{372}$$

or, equivalently,

$$S/N = \eta_c \frac{\pi}{8k} \frac{\sqrt{\eta_{A_1}\,\eta_{A_2}}\,L_1\,L_2\,\mathcal{S}_\nu}{\sqrt{T_{S_1}\,T_{S_2}}}\,\sqrt{2\,B\,\tau_a}, \tag{373}$$

where k, η_c, T_{S_1} and T_{S_2}, B, and τ_a are Boltzmann constant, the coherence factor, system noise temperatures, fequency bandwidth, and integration time, respectively, as before.

Furthermore, by introducing system equivalent flux densities (SEFDs) of the antennas,

$$SEFD_1 = \frac{2\,k\,T_{S_1}}{A_{e_1}}, \quad \text{and} \quad SEFD_2 = \frac{2\,k\,T_{S_2}}{A_{e_2}},$$

we have

$$S/N = \eta_c \frac{\mathcal{S}_\nu}{\sqrt{SEFD_1\;SEFD_2}}\,\sqrt{2\,B\,\tau_a}. \tag{374}$$

For example, if we observe a source with correlated flux density 0.2 Jy, with $L_1 = L_2 = 20$ m, $T_{S_1} = T_{S_2} = 120$ K, $\eta_{A_1} = \eta_{A_2} = 0.5$, $B = 256$ MHz, $\eta_c = 0.88$, and $\tau_a = 100$ sec, then we expect to obtain a signal–to–noise ratio $S/N \cong 19$, which is clearly detectable.

On the other hand, if we denote a limiting S/N value, necessary for detection of a source by an interferometer, as $(\mathcal{S}/\mathcal{N})$, we can derive from

equation (373) the minimum correlated flux density $\mathcal{S}_{\nu min}$ detectable by the interferometer:

$$\mathcal{S}_{\nu min} = (\mathcal{S}/\mathcal{N})\,\frac{8k}{\pi\eta_c}\,\frac{\sqrt{T_{S_1}\,T_{S_2}}}{\sqrt{\eta_{A_1}\,\eta_{A_2}}\,L_1\,L_2\,\sqrt{2\,B\,\tau_a}}. \tag{375}$$

For a VLBI, a signal–to–noise ratio of $6 \sim 7$ is usually adopted as the detection threshold $(\mathcal{S}/\mathcal{N})$. If we adopt $(\mathcal{S}/\mathcal{N}) = 7$, and assume again $L_1 = L_2 = 20$ m, $T_{S_1} = T_{S_2} = 120$ K, $\eta_{A_1} = \eta_{A_2} = 0.5$, $B = 256$ MHz, $\eta_c = 0.88$ and $\tau_a = 100$ sec, then we obtain $\mathcal{S}_{\nu min} = 0.074$ Jy.

We can observe many radio sources with such a sensitivity. In particular, almost all extragalactic continuum spectrum radio sources which are regularly used in international geodetic VLBI observations are readily detected, since their flux densities are mostly $\mathcal{S}_\nu > 0.1$ Jy.

4.3.7 Signal–to–Noise Ratio of a Single Dish Radio Telescope

Before closing this section, we will consider the signal–to–noise ratio formula for a single dish radio telescope which we introduced in Chapter 2:

$$S/N = \frac{T_A}{T_S}\sqrt{B\tau_a}, \tag{376}$$

where T_A, T_S, B, and τ_a are antenna temperature, system noise temperature, frequency bandwidth, and integration time, respectively, of the single dish radio telescope. This equation is very similar to equation (370) for an interferometer, except for the lack of an extra factor 2 in the square root of equation (376).

We can derive equation (376) on the basis of a statistical theory of the dispersion of a time–averaged squared Gaussian noise, which serves as a model for an output of a square–law detector (see, for example, Rolfs and Wilson, 2000). We will show here a variant of the derivation using a "correlator model" of a square–law detector as illustrated in Figure 85.

Usually, a correlator multiplies and then time–averages received voltages from two antennas of an interferometer (see Figure 83). However, we can use the same correlator as a square–law detector of a single dish telescope by feeding identical voltages from an antenna to the two inputs of the correlator (right pannel of Figure 85). In this sense, we can always use such a correlator as a model of a square–law detector.

Therefore, we can make use of all preceding discussions in this section for considering the signal–to–noise ratio of a single dish radio telescope if we put

$$U_1(t) = U_2(t) = U(t), \quad V_1(t) = V_2(t) = V(t), \quad N_1(t) = N_2(t) = N(t),$$

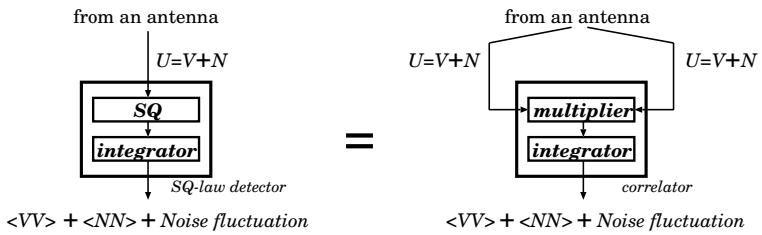

Figure 85: A square–law detector is equivalent to a correlator with identical inputs containing identical signals V and identical system noises N.

$$T_{A_1} = T_{A_2} = T_A, \;\; T_{S_1} = T_{S_2} = T_S, \;\; H_1(\omega) = H_2(\omega) = H(\omega),$$
$$G_1 = G_2 = G, \;\; \text{and} \;\; \phi_{LO_1} = \phi_{LO_2}.$$

However, there is an important difference between the interferometer and single dish cases.

In the single dish case, $\langle N_1(t)N_2(t)\rangle$ is no longer zero unlike in equation (314) for the interferometer case, since $N_1(t) = N_2(t) = N(t)$ and therefore

$$\langle N_1(t)N_2(t)\rangle = \langle N(t)N(t)\rangle \neq 0. \tag{377}$$

This shows the high "noise floor" in the square–law detector output of a single dish radio telescope, which does not exist in the correlator output of an interferometer.

Accordingly, we have to modify some of preceding equations. In particular, equations (315) and (316) must be replaced in the single dish case by equations:

$$\langle y(t)\rangle = \langle V(t)V(t)\rangle + \langle N(t)N(t)\rangle, \tag{378}$$

and

$$z(t) = \langle V(t)V(t)\rangle + \langle N(t)N(t)\rangle + \text{noise}. \tag{379}$$

Likewise, auto– and cross–correlations as well as power and cross–power spectra in equations (347) and (348) are now given by

$$R_{U_1U_1}(\tau) = R_{U_2U_2}(\tau) = R_{U_1U_2}(\tau) = R_{UU}(\tau) = R_{VV}(\tau) + R_{NN}(\tau), \tag{380}$$
$$S_{U_1U_1}(\omega) = S_{U_2U_2}(\omega) = S_{U_1U_2}(\omega) = S_{UU}(\omega) = S_{VV}(\omega) + S_{NN}(\omega). \tag{381}$$

Then, equations (353) and (355) are reduced to

$$S_{U_1U_2}(\omega) = S_{UU}(\omega) = \frac{1}{2}\, k\, [T_A(\omega_{LO}+\omega)+T_S(\omega_{LO}+\omega)] \;\mid H(\omega)\mid^2 \quad (\text{for } \omega \geq 0), \tag{382}$$

which is just the same with equation (352).

Therefore, equations (360), (361) and (362) for the dispersion of the "correlator" (square–law detector) output are modified to

$$\begin{aligned}\sigma_z^2 &= \frac{1}{2\pi^2} \int\limits_{-\infty}^{\infty} 2\, S_{UU}^2(\omega')\, d\omega' \int\limits_{-\infty}^{\infty} \mid \mathsf{A}(\omega) \mid^2 \, d\omega \\ &= \frac{k^2}{8\pi^2} \int\limits_{0}^{\infty} 2\,(T_A + T_S)^2 \, \mid H(\omega') \mid^4 \, d\omega' \int\limits_{-\infty}^{\infty} \mid \mathsf{A}(\omega) \mid^2 \, d\omega \\ &\cong \frac{k^2\, T_S^2}{4\pi^2} \int\limits_{0}^{\infty} \mid H(\omega') \mid^4 \, d\omega' \int\limits_{-\infty}^{\infty} \mid \mathsf{A}(\omega) \mid^2 \, d\omega. \end{aligned} \tag{383}$$

Now, in case of a rectangular filter as given by equation (365), we have

$$\mathcal{A} = k\, T_A\, G\, B, \tag{384}$$

for the signal part $\mathcal{A}$ in the "correlator" (square–law detector) output (see equations (363) and (366)), and

$$\int\limits_{0}^{\infty} \mid H(\omega') \mid^4 \, d\omega' = 2\pi\, G\, B, \tag{385}$$

(see equation (367)).

From equations (383), (384), (385), and (369), we obtain

$$S/N = \frac{\mathcal{A}}{\sigma_z} = \frac{T_A}{T_S}\sqrt{B\,\tau_a},$$

i.e. equation (376) for the signal–to–noise ratio of a single dish radio telescope.

It is now clear that the lack of a factor 2 in the square root of equation (376), which exists in equation (370), is explained by the fact that the output of a square–law detector has the high "noise floor" ($\langle NN \rangle \neq 0$), while the correlator output of an interferometer does not ($\langle N_1 N_2 \rangle = 0$).

References

Papoulis, A., 1984, Probability, Random Variables, and Stochastic Processes, 2nd Edition, *McGraw-Hill* (New York).

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P., 1988, Numerical Recipes in C, The Art of Scientific Computing, *Cambridge University Press* (New York).

Rohlfs, K., and Wilson, T. L., 2000, Tools of Radio Astronomy, 3rd Edition, *Springer–Verlag* (Berlin).

Thompson, A.R., Moran, J.M., and Swenson, G.W., Jr., 2001, Interferometry and Synthesis in Radio Astronomy, *John Wiley & Sons* (New York).