Very Long Baseline Interferometry and the VLBA
ASP Conference Series, Vol. 82, 1995
J. A. Zensus, P. J. Diamond, and P. J. Napier (eds.)

Part I

Basic Theory

Very Long Baseline Interferometry and the VLBA
ASP Conference Series, Vol. 82, 1995
J. A. Zensus, P. J. Diamond, and P. J. Napier (eds.)

CHAPTER 1

Interferometers and Coherence Theory

B. G. CLARK

NATIONAL RADIO ASTRONOMY OBSERVATORY

Abstract A brief overview of the fundamental principles underlying radio interferometry is presented using the terminology of modern optics . This includes the fundamental equations of aperture synthesis, a definition of terms for interferometers, and a brief discussion of the problem of calibrating interferometer phase.

Contents

1.1 Introduction—Basis and Textbooks

This lecture comprises a quick survey of the principles of synthesis imaging, and a brief discussion of the theoretical underpinning of the special problems and techniques encountered in imaging in long baseline interferometry. This will be carried out in the terminology of modern optics. Until recently, it was the custom in radio astronomy to discuss interferometry in the terminology of the electrical engineering discipline of the founders of the technique, so those with some familiarity with the field may find this lecture more obscure than they were expecting.

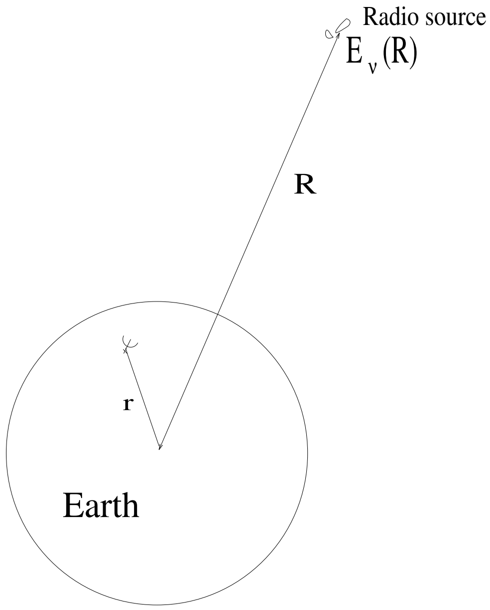

Figure 1.1: Radio source–antenna geometry

Still the best reference for the basics of modern optics is the excellent textbook *Principles of Optics*, by Born and Wolf (1980). Their chapter X is especially relevant to this summer school. For a more particularized discussion of radio interferometers, see *Interferometry and Synthesis in Radio Astronomy*, by Thompson, Moran and Swenson (1986). A broader review of radio imaging and radio telescopes is given by Christiansen and Högbom (the second, 1985, edition is recommended). A slightly different viewpoint (that of stochastic signal processing) is given in *Statistical Optics* by Goodman (1985).

1.2 Electromagnetic Field and Field Propagation

I will start with a simple derivation of the fundamental equations of aperture synthesis. A number of simplifying assumptions will be introduced which are valid to a high order for the practical cases of special interest to us in VLBI observations of astrophysical radio sources.

We consider the case of a very distant celestial radio source, located at a position $\mathbf{R}$ (boldface symbols indicate vectors, in this case a position vector). This object is radiating, producing a time variable electric field over the face of the source, $\mathbf{E}(\mathbf{R}, t)$. Maxwell's equations say that an electromagnetic wave will propagate away from the source and will eventually arrive at a point where an astronomer may conveniently observe it, say at the point $\mathbf{r}$. See figure 1.1.

It is convenient to consider the frequency components of the time-varying field. For a finite time interval, we may express the electric field as the real part of the sum of a Fourier series, in which the only time-varying functions are simple exponentials, and, hereafter, deal only with the coefficients of this Fourier series, rather than with a general time variation. The coefficients of this Fourier series, which I shall call $\mathbf{E}_\nu(\mathbf{R})$, are called the *quasi-monochromatic components* of the electric field $\mathbf{E}(\mathbf{R}, t)$. Note that the fields $\mathbf{E}_\nu(\mathbf{R})$ are complex quantities, and it is useful to think of them as such at all times. It leads to a more elegant formulation of the theories to consider this complex nature to be physical reality rather than a mathematical convenience.

For now, let us consider only a single quasi-monochromatic component. Later we shall consider the total response, as a sum of the responses to all the components. The linearity property of Maxwell's equations allows us to superpose the fields at our observer location $\mathbf{r}$ by the various source points,

$$\mathbf{E}_\nu(\mathbf{r}) = \int P_\nu(\mathbf{R}, \mathbf{r}) \mathbf{E}_\nu(\mathbf{R})\, dS\,. \tag{1.1}$$

The integral is to be taken over the radio source. The function $P_\nu(\mathbf{R}, \mathbf{r})$ is called the *propagator*, and describes how the electric field at $\mathbf{R}$ influences the electric field at $\mathbf{r}$. It is, properly, a tensor function, describing how each component of the vector field at the source influences each component of the field at the point of reception. But, at this point, we introduce the first of our simplifying assumptions, that P is an ordinary scalar function. For this to hold, it suffices (though is not required) that the space between us and the radio source be empty. And for empty space, the propagator takes a particularly simple form:

$$\mathbf{E}_\nu(\mathbf{r}) = \int \frac{\mathbf{E}(\mathbf{R}) e^{2\pi i \nu |\mathbf{R} - \mathbf{r}|/c}}{|\mathbf{R} - \mathbf{r}|}\, dS\,. \tag{1.2}$$

This equation is the general form of the quasi-monochromatic component of the electric field at frequency ν due to the radiation from the cosmic radio source. With few exceptions, this is all we have in astronomy; we can measure only the properties of this field $\mathbf{E}_\nu(\mathbf{r})$ to tell us about the nature of things in the universe.

1.3 The Spatial Coherence Function of the Field

Among the properties of $\mathbf{E}_\nu(\mathbf{r})$ is the correlation of the field at two different locations. The correlation of the field at points $\mathbf{r}_1$ and $\mathbf{r}_2$ is defined as the expectation of the product:

$$\mathbf{V}_\nu(\mathbf{r}_1, \mathbf{r}_2) = \langle \mathbf{E}_\nu(\mathbf{r}_1) \mathbf{E}^*_\nu(\mathbf{r}_2) \rangle\,. \tag{1.3}$$

The raised asterisk indicates the complex conjugate. We can then use equation 1.2 to substitute for $\mathbf{E}_\nu(\mathbf{r})$, and write the product of the integrals as a double integral over two separate surface element dummy variables:

$$\mathbf{V}_\nu(\mathbf{r}_1, \mathbf{r}_2) = \left\langle \iint \mathbf{E}_\nu(\mathbf{R}_1)\mathbf{E}_\nu(\mathbf{R}_2) \frac{e^{2\pi i\nu|\mathbf{R}_1-\mathbf{r}_1|/c}}{|\mathbf{R}_1-\mathbf{r}_1|} \frac{e^{-2\pi i\nu|\mathbf{R}_2-\mathbf{r}_2|/c}}{|\mathbf{R}_2-\mathbf{r}_2|} \, dS_1 \, dS_2 \right\rangle . \tag{1.4}$$

This expression simplifies under the very good assumption that the radiation from astronomical objects is not spatially coherent; that is, that $\langle \mathbf{E}_\nu(\mathbf{R}_1)\mathbf{E}_\nu(\mathbf{R}_2)\rangle$ is zero for $\mathbf{R}_1 \neq \mathbf{R}_2$. Exchanging the expectation and integral in equation 1.4 then gives

$$\mathbf{V}_\nu(\mathbf{r}_1, \mathbf{r}_2) = \int \left\langle |\mathbf{E}(\mathbf{R})|^2 \right\rangle |\mathbf{R}|^2 \frac{e^{2\pi i\nu|\mathbf{R}-\mathbf{r}_1|/c}}{|\mathbf{R}-\mathbf{r}_1|} \frac{e^{-2\pi i\nu|\mathbf{R}-\mathbf{r}_2|/c}}{|\mathbf{R}-\mathbf{r}_2|} \, dS \, . \tag{1.5}$$

This can be further simplified by noting the great distance to cosmic radio sources, which means that in the cases of current interest, $|\mathbf{R}|$ is of order parsecs to megaparsecs, whereas the vector $\mathbf{r}$, measuring the distance to different sites of observation, is of order tens of meters to thousands of kilometers. We can, to a high order of accuracy, ignore terms of order $|\mathbf{r}|/|\mathbf{R}|$, in particular in the denominators of the equation above.

$$|\mathbf{R} - \mathbf{r}| = \sqrt{|\mathbf{R}|^2 + |\mathbf{r}|^2 - 2\mathbf{r}\cdot\mathbf{R}} \approx |\mathbf{R}| - \frac{\mathbf{r}\cdot\mathbf{R}}{|\mathbf{R}|} \, .$$

If, for the moment, we ignore the vector nature of equation 1.5, we can cast it into terms of the observed *intensity* of the radiation field, $I = |\mathbf{R}|^2 \langle|\mathbf{E}_\nu|\rangle^2$. We will also use a unit vector, $\mathbf{s}$, in the direction of $\mathbf{R}$ (that is, $\mathbf{s} = \mathbf{R}/|\mathbf{R}|$), and replace the element of integration, dS, with the element of solid angle, $d\Omega$. These substitutions yield:

$$V_\nu(\mathbf{r}_1, \mathbf{r}_2) = \int I_\nu(\mathbf{s}) e^{-2\pi i\nu \mathbf{s}\cdot(\mathbf{r}_1-\mathbf{r}_2)/c} \, d\Omega \, . \tag{1.6}$$

Note that equation 1.6 depends only on the separation vector $\mathbf{r}_1 - \mathbf{r}_2$, not on the absolute locations, $\mathbf{r}_1$ and $\mathbf{r}_2$. Therefore, we can find out all we can learn about the correlation properties of the radiation field by holding one observation point fixed, and moving the second around; we do not have to measure at all possible pairs of points. This function, V_ν, of the single (vector) separation $\mathbf{r}_1 - \mathbf{r}_2$ is called the *spatial coherence function*, or the *spatial autocorrelation function*, of the field $\mathbf{E}_\nu(\mathbf{r})$. It is the simplest property of the radiation field that we can measure.

An interferometer is a device for measuring this spatial coherence function.

1.4 Synthesis Imaging

Equation 1.6 is, within reasonable, well-defined limits, invertible. The intensity distribution of the radiation as a function of direction $\mathbf{s}$ can therefore be deduced in certain cases by measuring the spatial coherence function V as a function of $\mathbf{r}_1 - \mathbf{r}_2$ and performing the inversion.

To make an explicit inversion of this equation, we can make a further simplifying assumption, which holds very well for those cases toward which this summer school is directed. This is that the source is of small angular size; that is, that the vector $\mathbf{s}$ may be expanded, over the radio source, as $\mathbf{s} = \mathbf{s}_0 + \sigma$, in which we may neglect all terms in the squares of the components of σ. Since $\mathbf{s}$ and $\mathbf{s}_0$ are, by definition, unit vectors, we have

$$\mathbf{s} \cdot \mathbf{s} = \mathbf{s}_0 \cdot \mathbf{s}_0 = 1$$

$$\mathbf{s} \cdot \mathbf{s} = \mathbf{s}_0 \cdot \mathbf{s}_0 + 2\mathbf{s}_0 \cdot \sigma + \sigma \cdot \sigma \approx 1 + 2\mathbf{s}_0 \cdot \sigma\,.$$

Since their dot product is zero, $\mathbf{s}_0$ and σ are perpendicular. We can introduce a special coordinate system such that $\mathbf{s}_0 = (0, 0, 1)$. In this coordinate system,

$$\mathbf{r}_1 - \mathbf{r}_2 = c(u, v, w)/\nu\,,$$

and, to this approximation, $\mathbf{s} = (x, y, 1)$. With the above approximation and this coordinate system, equation 1.6 on the facing page becomes

$$V_\nu(u, v, w) = e^{-2\pi i w} \iint I_\nu(x, y) e^{-2\pi i(ux+vy)}\, dx\, dy\,. \tag{1.7}$$

It is customary to absorb the factor in front of the integral into the left hand side of this equation. At the risk of causing confusion, I shall use the same symbol, V, with two arguments rather than three, for doing this:

$$V_\nu(u, v) = \iint I_\nu(x, y) e^{-2\pi i(ux+vy)}\, dx\, dy\,. \tag{1.8}$$

$V_\nu(u, v)$ is the coherence function relative to the direction $\mathbf{s}_0$, which is called the *phase tracking center.*

Since equation 1.8 is a Fourier transform, it may be formally inverted:

$$I_\nu(x, y) = \iint V_\nu(u, v) e^{2\pi i(ux+vy)}\, du\, dv\,. \tag{1.9}$$

In practice, the spatial coherence function V is not known everywhere, but is sampled at particular places on the u-v plane. The sampling can be described by a *sampling function* $S(u, v)$, which is zero where no data have been taken. Then, instead of equation 1.9, one may actually calculate

$$I_\nu^D(x, y) = \iint V_\nu(u, v) S(u, v) e^{2\pi i(ux+vy)}\, du\, dv\,. \tag{1.10}$$

Radio astronomers often refer to $I_\nu^D(x, y)$ calculated this way as the *dirty image*; its relation to the desired intensity distribution $I_\nu(x, y)$ is seen by using the Fourier convolution theorem:

$$I_\nu^D = I_\nu * B\,, \tag{1.11}$$

where the asterisk denotes convolution, and

$$B(x, y) = \iint S(u, v) e^{2\pi i(ux+vy)}\, du\, dv \tag{1.12}$$

is called the *synthesized beam*, *point spread function*, or *dirty beam* corresponding to the sampling function $S(u, v)$. Equation 1.11 on the previous page says that I^D is the true intensity distribution I convolved with the synthesized beam B.

1.5 Spectroscopy

The discussion above has been based on the "quasi-monochromatic components" introduced at the very beginning. This concept can, in fact, be physically implemented, by inserting filters in the interferometer signal paths, so that sufficiently (in a sense to be defined below) narrow band "quasi-monochromatic components" are individually handled. An individual quasi-monochromatic component carries only a tiny fraction of the information present in the output of the interferometer front-end amplifiers, so if this is done in practice, many quasi-monochromatic components are extracted from the signal in parallel, and processed simultaneously. (In practice, the multi-filter approach may be implemented by either electronic components, or by digital filtering, especially by a fast Fourier transform.) In this case, then, we can regard each quasi-monochromatic component as carrying information about the spectrum of the intensity I_ν at its frequency, and the problem is not one of understanding how interferometric spectroscopy works, but of understanding how spectral channels can be added together, to perform a continuum measurement of the source brightness.

At this point, we step aside to ask the question of how restricted in frequency a quasi-monochromatic component must be. The answer is readily apparent from equation 1.8. The range of frequencies must be sufficiently small that the right hand side does not appreciably change over it. For this, it suffices if (a) The intensity I_ν does not appreciably change over the range, and (b) $(ux + vy)$ does not appreciably change over the range. The latter changes because of the term ν in the definition of u and v. For this change to be small, it suffices that the source size, θ_{max}, times the maximum baseline of interest, should be small compared to one:

$$\theta_{max} |\mathbf{r}| \nu / c \ll 1 \, .$$

This condition is often met over the entire source of interest, and, for continuum emission for which I_ν changes very slowly, the treatment above for quasi-monochromatic components carries over without change into the continuum intensity.

One might think that one also has to take into account the variation in the leading term of equation 1.7, $e^{-2\pi i w}$, since this also varies proportionately to ν. And indeed one would have to, if the electronic equipment responsible for converting from equation 1.7 to equation 1.8 can only calculate the center frequency value of this term. However, it is fairly easy to implement, in electronics, a term that varies proportionately to frequency. Such a term is called a delay.

In fact, a similar viewpoint taken of equation 1.7, reproduced below, presents an alternate to the multi-filter approach to performing spectroscopy on I_ν described above.

$$V_\nu(u, v, w) = e^{-2\pi i w} V_\nu(u, v) . \tag{1.13}$$

If we replace w (the naturally occurring component of the interferometer baseline along the direction to the source, with $w + \tau\nu$, where τ is an additional, electronically induced delay (which is, remember, a phase rotation proportional to frequency), and then add together all the quasi-monochromatic components, we have (again threatening confusion by reusing the same symbol V),

$$V(u, v, \tau) = \int e^{-2\pi i \tau\nu} V_\nu(u, v)\, d\nu . \tag{1.14}$$

This is patently the Fourier transform of V_ν, and may be used instead of the multi-filter approach. The electronic implementation of the correlation function calculation is often simpler than implementation of multiple filters.

The first approach is called the FX approach (frequency domain cross-multiplication), the latter the XF approach (cross-multiplication, Fourier transform) or simply the *lag correlation* approach.

1.6 Polarization Imaging

At this point we acknowledge that we have slipped in an extra assumption. We moved a little more quickly than we were entitled to from talking about the electric field values $\mathbf{E}_\nu(\mathbf{R})$ to the intensity, $I_\nu(\mathbf{R})$. And we can also now mention why, to this point, we have been careful to say *quasi-monochromatic component*, rather than dealing with truly monochromatic components. For a monochromatic electromagnetic field, the electric field is described throughout space and time by a single vector quantity. The field is, in some sense, completely polarized, since the field variation in any vector direction is completely described by that in the polarization sense reference. For the noise-like emission of interest to radio astronomers, though, polarization is defined by a correlation process. One picks two orthogonal polarizations and analyses the radiation of the quasi-monochromatic fields into the components in the two polarizations. The polarization of the quasi-monochromatic wave is described by the 2×2 matrix of correlations between these two resolutions into orthogonal components. For instance, if we pick right and left circular polarization as the two orthogonal modes (the discussion is very little changed if we pick two orthogonal linears), then the matrix

$$\begin{pmatrix} \langle RR^* \rangle & \langle RL^* \rangle \\ \langle LR^* \rangle & \langle LL^* \rangle \end{pmatrix} \tag{1.15}$$

describes the polarization. This can be related to more familiar descriptions of polarization. For instance, the *Stokes parameters* have the intensity I, two linear polarizations parameters Q and U, and a circular polarization parameter V, related to the above numbers in simple (and more or less obvious) linear combinations:

$$\begin{pmatrix} \langle RR^* \rangle & \langle RL^* \rangle \\ \langle LR^* \rangle & \langle LL^* \rangle \end{pmatrix} = \begin{pmatrix} I+V & Q+iU \\ Q-iU & I-V \end{pmatrix} . \tag{1.16}$$

The complex correlation functions at the celestial radio source are preserved in the spatial coherence functions that interferometers measure, under the assumption that space between the interferometer and the source is empty. That is, one can derive, for instance, the distribution of $\langle RR^* \rangle$ on the radio source by measuring the coherence function of R on the ground, and so forth for the other components of the matrix above. Since the intensity is the quantity in which one is always interested, one usually forms the sum of the L and R coherence functions before transforming to the sky plane, which one can always do, since the relations are linear. One can do the same with the other Stokes parameters, or the $\langle RL^* \rangle$, etc., can be transformed directly, and the resulting distribution of $\langle RL^* \rangle$, etc., on the sky can be used to convert to the Stokes parameters.

For much of the work in VLBI, it has been conventional to assume that the circular polarization of radio sources is zero, and to present an image of $\langle LL^* \rangle$, for example, as an image of intensity. For most cases of interest, this probably introduces little error, but there are times one must bear in mind the difference.

1.7 Design of Radio Interferometers

Almost all radio interferometers have the generic block diagram shown in figure 1.2 on the facing page. This shows a simple two-element interferometer. It can easily be extended to multi-element arrays by replicating the element-based portion, and considering that the final portion, the correlator which compares the signals from the two elements, must have a portion for each *pair* of elements. (For a multi-element array of N elements, there are $N(N-1)/2$ pairs, or *baselines*.)

At each interferometer element (sometimes called *station*, or just *antenna*), there is an antenna system, which collects the electromagnetic energy over a fairly large area (usually hundreds of square meters) and feeds it into a single waveguide or coaxial cable. Then there is a low-noise front end amplifier, which amplifies the excessively weak celestial radiation (while contributing its own noise) to levels easier to deal with. (Some radio interferometers omit this stage—their first active electronic device is a mixer.) For the purposes of figure 1.2, this low-noise front end amplifier is considered to be part of the antenna system.

The next element in the design is the local oscillator and mixer system. The goal of the local oscillator system is to produce an electric field varying as a

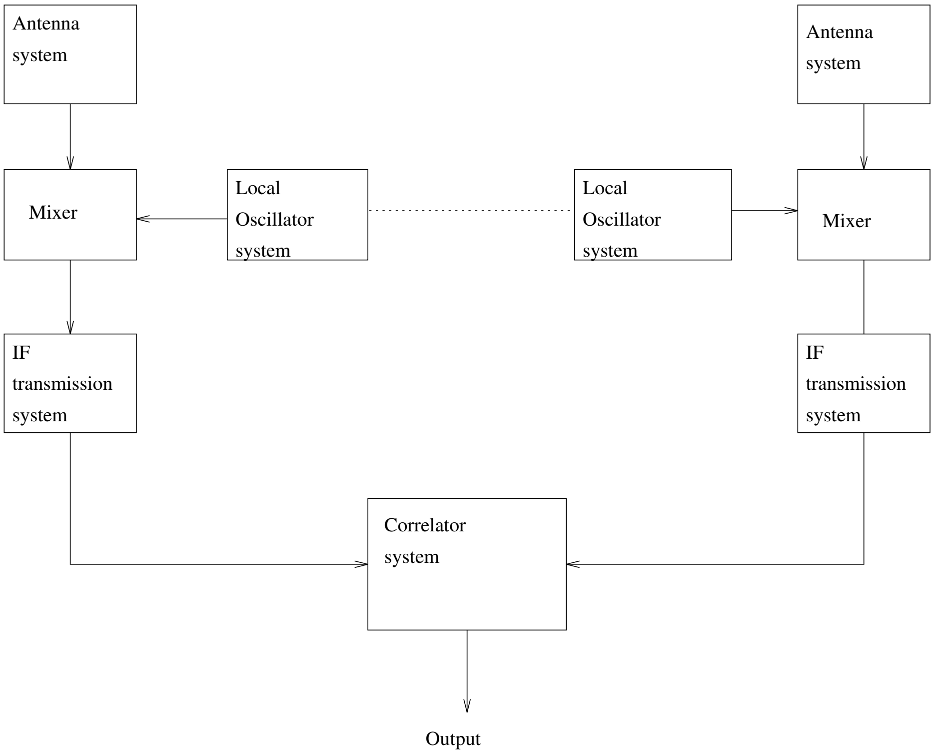

Figure 1.2: Generic interferometer block diagram

pure sine wave, identical at each element of the interferometer. This can be done in several ways. The simplest is to generate a sine wave at some central location, and sending it to each element over identical, stable, paths. For long baseline interferometers, this has severe complications. What is usually done is to produce as pure and stable a sine wave possible at each element, and rely on the generic similarity of sine waves. The art of producing high quality oscillators has progressed rapidly since the first introduction of atomic transition controlled oscillators about thirty-five years ago, and, with the very best currently available oscillators, the quality of the local oscillator system is no longer a limiting factor.

The next element in the design is the IF transmission system. The goal of this system is to transfer the celestial signals, now converted to a lower frequency for convenience in processing, from each interferometer element to a central location where the signals may be correlated. It must do so with fidelity, and must control the delay in the system so that samples of the same wavefront of radiation from the celestial source arrive at the correlator at the same time. Many options are employed for this system. For short distances, coaxial cables are suitable. For slightly longer distances, fiber optic cables are suitable. (The

VLA is connected by waveguide, a once promising communication technology which has been superseded by fiber optics, which have similar capabilities at much lower cost.) For long baseline interferometers, the design of choice is recording on magnetic tape, and transporting the IFs by planes, trains, and automobiles.

The final element in this block diagram is the correlator. In the simplest case it is a simple multiply-and-average device. It may also include, as discussed above, filter banks or correlation at many delays, for spectroscopic interferometry.

1.8 Element Gain

At this point, it is necessary to bring our abstract, aesthetically pleasing, description of aperture synthesis somewhat nearer to the way things happen in the real world. We said above simply, "An interferometer is a device for measuring this spatial coherence function." However, interferometers have characteristics of their own, which, in practicality, we must acknowledge. To a reasonable approximation, the personality of the interferometer may be characterized by a single complex number for each element, the *element gain*.

That is, the quantity measured is not the true coherence function $V_\nu(u,v)$, but is $G_{1,\nu}G^*_{2,\nu}V_\nu(u,v)$, where the element gains of the two elements, $G_{1,\nu}$ and $G_{2,\nu}$ are different for the two elements, and are, unfortunately, time variable.

Determination of the element gains is the place that real difficulty enters the determination of brightness distributions by interferometry. As I shall discuss below, simple self-contained determinations of the gains, especially the phase of the complex gain, is often not possible, and external procedures must be used. Adding to the complexity is the fact that these phases may be varying on time scales of minutes, or, at the shorter wavelengths, seconds.

One source of this phase variation is the fact that the phase of the sine wave delivered by the local oscillator system differs from the desired perfect sine wave. This has historically been the primary source of phase uncertainty; only recently have oscillators become good enough that other effects (primarily the atmosphere–see below) often dominate the phase uncertainty.

It is also convenient to lump into the element gain the effects of the earth's atmosphere on the transmitted signal. This also can be characterized by a frequency dependent complex number. The magnitude of this number is caused by absorption (except at very low frequencies, where more complex phenomena are seen). At microwave frequencies this arises mainly by absorption by water vapor. At frequencies above about 30 GHz, absorption by oxygen molecules becomes important. At the higher frequencies, also, absorption by liquid water is important if clouds are present. The phase portion of this number arises from the refractivity differing from unity. At low frequencies, the refractivity of the ionosphere introduces phase inversely proportional to frequency, which varies on diurnal, minutes (ionospheric traveling disturbances), and occasionally, seconds time scales. At microwave frequencies, the troposphere introduces variations proportional to frequency (that is, delays). These are due to the refractivity introduced by oxygen and nitrogen, and that due to water vapor. The latter is of disproportionate importance because of the high dipole moment of the water molecule; a gram per square centimeter of water above an element will introduce several times as much delay as a gram per square centimeter of dry atmosphere. It is even more important when considering time variations; turbulent variations of water vapor content are long-lived and important, whereas density variations in the dry component of the atmosphere large enough to be important for radio wavelengths are not.

An additional term that may complicate things arises from the fact that the locations of the elements, $\mathbf{r}$, are not a priori known with infinite precision. The elements are attached to the solid earth, whose quotidian deformations are pretty well known, but whose motions, in an inertial coordinate system are not. At a given instant, the angle through which the earth has rotated is uncertain by an amount corresponding to five cm or so in the location of a point on its surface. This is not really an element gain, but for some purposes it can be absorbed into this slowly varying term and otherwise ignored. Historically the motions of the earth were not very well modeled in VLBI observations. This was partially due to the fact that, phase variations being dominated by local oscillator imperfection, it did not seem worth doing, and partly because tracking the motion of the earth has become much more accurate in recent years, in large part due to the advances in long baseline interferometry and related techniques.

1.9 Calibrating Element Gains

In practice, calibrating the magnitudes of the gains and the phases of the gains are attacked by rather different techniques. Most of what I shall say below applies to calibration of the phases, which is an intrinsically more difficult problem than the calibration of the magnitudes. Most interferometer elements contain an automatic level control early in the IF transmission system. Since the total signal level is controlled at this point, the problem of calibrating the magnitude of the gain reduces to the problem of estimating the proportion of this total signal level provided by the celestial radio source. There are a number of adequate techniques for doing this which are self-contained within the environment of the interferometer element (all variants of merely noting how much signal levels change when the element is pointed away from the source of interest). These methods require great care in execution and advance planning to make sure that the required data are collected at the appropriate times, but the techniques involved are insufficiently colorful to attract our attention at this time.

In connected interferometers the practice is to calibrate element gains by interspersing observations of a point source of known position, called a *calibrator*, among the observations of the unknown source being observed. For a point, or unresolved, source, the visibility is a known constant: $V = V_0$. Therefore, for an interferometer array, we have measurements $A_{i,j}$ for each baseline (at this point, for simplicity, we drop the ν subscript):

$$A_{i,j} = G_i G_j^* V_0 \,. \tag{1.17}$$

This is clearly a system of $N(N-1)/2$ equations in N unknowns, and can be solved for the gains G for $N > 2$. There is one minor exception. Since the gains appear only in $G_i G_j^*$ pairs, any solution to the equations, the set G_i, remains a solution if all members are multiplied by a number of unit magnitude and arbitrary phase. This means we may arbitrarily assign one phase, and solve for the rest. It is conventional to assign some element (usually called *the reference antenna*) a zero phase. (An alternate is to require that the sum of all phases is zero.) Although the choice of phase is entirely arbitrary at any moment, one uses these instantaneously determined gains to interpolate to the appropriate gains for the time of the observation of the unknown; this process of interpolation works best if the reference antenna chosen is itself one of the best behaved of the array.

This procedure can also be applied in long baseline interferometry. It is technically somewhat more difficult because 1) on baselines comparable to the radius of the earth, there are no point sources, all show some structure; 2) the position of the calibrator is often not known with sufficient accuracy; and 3) there are various important effects of long scale length which can be neglected for shorter interferometers because they contribute the same phase to all elements. Examples of the last are ionospheric structures (hundred-kilometer waves, the equatorial electrojet phenomenon, and the very strong diurnal effect), and the overall phase path due to the troposphere. These tend to place strong restrictions on the distance between the source and calibrator, which are much less severe for shorter interferometers.

This theoretically straightforward but technically difficult calibration of the gains is a fairly recent development in long baseline interferometry. Because other methods of phase estimation were in use at the time, the procedure is not called simply *phase calibration* as it is for the connected interferometers, but *phase referencing* or *phase connection* (the phase of the unknown is considered referenced or connected to that of the calibrator).

1.10 Living with Uncalibrated Gains

The earliest long baseline interferometry was done by giving up all hope of measuring phases and fitting simple models to the amplitudes. This seems to work well in practice, but is somewhat dangerous. The model fitting routines will give the best fitting model of the sort proposed, but cannot tell the user if it is the most appropriate sort of model. In fact, one usually accepts models that formal statistical techniques assert are significantly different from the data. It is considered more dangerous to elaborate the model, introducing more parameters, to get a better fit to the data than it is to accept an image that clearly does not describe the real source.

There is more information in the data from a multi-element interferometer than simply the amplitudes, even if the phases of the element gains are completely unknown. The non-point source equivalent of equation 1.17 on the preceding page is

$$A_{i,j} = G_i G_j^* V(u_{i,j}, v_{i,j}) . \tag{1.18}$$

We can, for instance, with a three (or more) element interferometer, form the product

$$A_{i,j} A_{j,k} A_{k,i} = |G_i|^2 |G_j|^2 |G_k|^2 V(u_{i,j}, v_{i,j}) V(u_{j,k}, v_{j,k}) V(u_{k,i}, v_{k,i}) , \tag{1.19}$$

which is independent of the phases of the G_i, and contains information about the brightness distribution of the source. Unlike the case of known gains, in which the visibilities are inverted by a Fourier transform to an image, there are no known formal methods for directly converting such things into images. The various procedures for utilizing this information rely on iterative inversions of one sort or another, comparing a starting image with the data, either in the form of equation 1.18 or in the form of equation 1.19, and from that comparison suggesting alterations to the starting image.

One procedure is to use the starting image to calculate a starting value for $V(u, v)$, to insert this into equations 1.18, to apply a least squares, maximum likelihood, or other estimation technique to solve the equations for the G_i, then to invert the equation to find the new $V(u, v)$. This is done for each set of samples the visibilities $A_{i,j}$, taken at different times, for the source. These are then Fourier inverted to make an image, and the image is then simplified (by means of the CLEAN algorithm or a variational method), and used as the starting image for the next step of the iteration. This procedure is called *self-calibration* if the method used to solve equation 1.18 for the G_i is to minimize the sum square error in the $A_{i,j}$, and *hybrid mapping* if the method is minimizing the errors in the phases of the $A_{i,j}$. (It has never been clear to me of what two things the method is a hybrid.)

If least squares is used to collapse the six-dimensional equation 1.19 onto a two-dimensional plane, and the image formed from that, the method is called the bispectrum method. This method has advantages for moderately weak sources, as the triple product can be integrated over times long compared to the time scale of changes in the G_i. In the other methods the integration must be stopped on a time scale such that the variations in the phase of the G_i are small compared to a radian.

1.11 Conclusion

All of the above has been merely a brief, non-technical, overview of the field. Many of the things mentioned above, when examined in detail, are replete with technical difficulties of their own, some to the extent that they approach an art, rather than being a simple recipe that can be followed from observation to image. The elaboration of these technicalities is the purpose of this summer school, to enable you to expand these elegant generalities into a practical tool for investigating the cosmos.

References

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

Christiansen, W. N., & Högbom, J. A. 1985. *Radiotelescopes* (Second ed.). Cambridge: Cambridge University Press.

Goodman, J. W. 1985. *Statistical Optics*. New York: John Wiley and Sons.

Thompson, A. R., Moran, J. M., & Swenson, G. W., Jr. 1986. *Interferometry and Synthesis in Radio Astronomy*. New York: Wiley-Interscience. First (1991) and second (1994) reprintings by Krieger Pub. Co., Malabar (Florida).