Synthesis Imaging in Radio Astronomy II
ASP Conference Series, Vol. 180, 1999
G. B. Taylor, C. L. Carilli, and R. A. Perley (eds.)

1. Coherence in Radio Astronomy

B. G. Clark
National Radio Astronomy Observatory, Socorro, NM 87801, U.S.A.

Abstract.
In this lecture the main principles of synthesis imaging are derived.

1. Introduction

It is appropriate for this specialized summer school to start with a survey of the derivation of the main principles of synthesis imaging, paying particular attention to the assumptions that go into them. This is because a substantial number of the lectures to follow will discuss the problems which arise when these assumptions are violated under the conditions of the observation the astronomer wants to make. At the same time, I will cast this introduction into the terminology of modern optics, in an attempt to stay abreast of current fashions in physics.

The fundamental reference for the basics of modern optics is the excellent textbook *Principles of Optics*, by Born and Wolf; their Chapter X is especially relevant to this summer school; another excellent reference on physical optics is *Statistical Optics* by Goodman. An excellent discussion of synthesis imaging, employing this modern terminology, is given by J. L. Yen (1985) in Chapter 5 of *Array Signal Processing* (S. Haykin, Ed.). A broader view of radio telescopes, again from a viewpoint of Fourier optics, but taking a somewhat historical perspective, is given in *Radiotelescopes* by Christiansen & Högbom (1985, Second Edition); their Chapter 7 discusses synthesis methods. The alternate viewpoint on radio interferometers, from the perspective of the electrical engineers who originally developed them, is explicated in Swenson & Mathur (1968).

2. Form of the Observed Electric Field

I will start with the most general formulation of the subject, and, one by one, introduce the simplifying assumptions until reaching the simple, elegant theoretical basis that is, after all, sufficient for much of radio interferometry. In the most general case, an astrophysical phenomenon occurs at location $\mathbf{R}$ (the boldface symbols indicate vectors, in this case a position vector). This phenomenon causes a time-variable electric field, which I will denote as $\mathbf{E}(\mathbf{R}, t)$. Then, Maxwell's equations say that an electromagnetic wave will propagate away from that point and will eventually arrive at a point where an astronomer may conveniently observe it, say the point $\mathbf{r}$. See Figure 1–1.

It is inconvenient for a number of reasons to deal with general time-variable electric fields. If we have a finite time interval of a varying field, we may express the magnitude of the field as the real part of the sum of the Fourier series in which the only time-varying functions are simple exponentials. Because of the linearity of Maxwell's equations (in the cases of astrophysical interest, anyway) we may deal with the coefficients of this Fourier series, rather than with the general time-varying field. The coefficients of this Fourier series, which I will

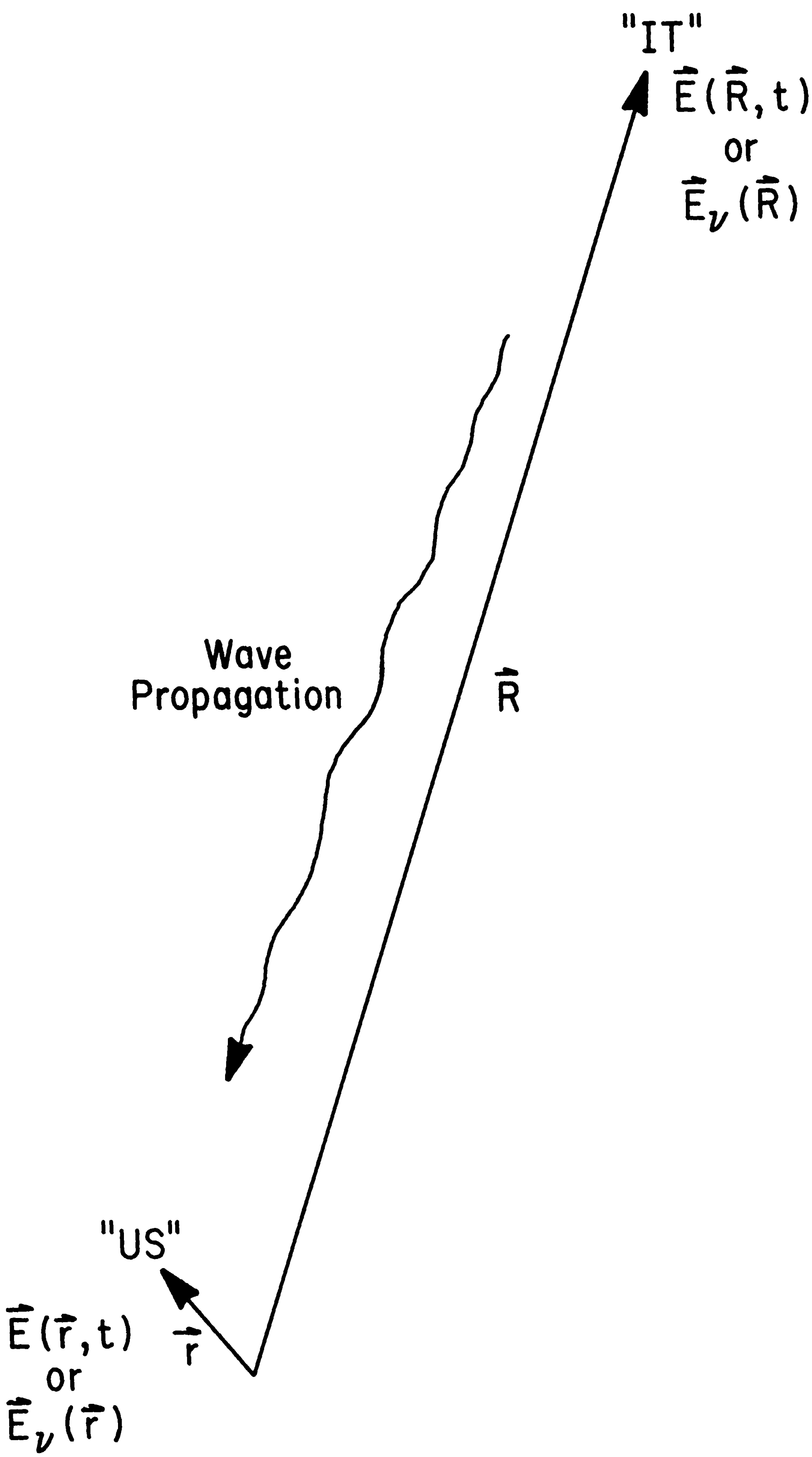

Figure 1–1. An astrophysically interesting phenomenon "IT", located at $\mathbf{R}$, causes an electromagnetic wave $\mathbf{E}_\nu(\mathbf{r})$ to propagate through space, where it can be detected at $\mathbf{r}$ by "US".

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.

In what follows, I consider only a single quasi-monochromatic component, realizing that the total response is the sum of the responses to all the components. In fact, the response of an instrument can be made arbitrarily close to that of a single quasi-monochromatic component, by inserting filters in the early, linear, parts of the instrument.

The linearity property of Maxwell's equations allows us to superpose the fields produced at a test location by the various source points,

$$\mathbf{E}_\nu(\mathbf{r}) = \iiint P_\nu(\mathbf{R}, \mathbf{r}) \mathbf{E}_\nu(\mathbf{R}) \, dx \, dy \, dz \, . \tag{1–1}$$

The integral is to be taken over all of space. 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 $\mathbf{r}$.

At this point, I begin to introduce the simplifying assumptions. The first assumption may be considered to be merely pedagogical, in the sense that it is not really needed at all, and is made only to avoid complicating the equations to the point that their physical meaning is obscured. For the moment, I shall ignore the fact that electromagnetic radiation is a vector phenomenon, and treat it as if it were simply a scalar field, measured at any point by a scalar quantity E. That is to say, I shall ignore, for the moment, all polarization phenomena. This enables the multiplication in Equation 1–1 to be regarded as ordinary scalar multiplication, and the propagator P to be an ordinary scalar function (not a tensor function as a complete derivation would have it).

The second simplifying assumption is that the sources of interest to astronomers are a long way away. The practical implication of this is that we have to give up all hope of describing the structure of the emitting regions in the third dimension, in depth. All we may measure is the "surface brightness" of the emitting phenomenon. One convenient way of expressing this assumption is to replace the field strength E at the source with the field strength at a convenient point distant from both us and from any source of radiation. We may conceive of a "celestial sphere", a very large sphere of radius $|\mathbf{R}|$, within which there is no additional radiation, and all that we may learn about the distribution of the source of the fields is the distribution of the electric field on the surface of this sphere, which I will call $\mathcal{E}_\nu(\mathbf{R})$. See Figure 1–2.

The third simplifying assumption is that the space within the "celestial sphere" is empty. For this case, Huygens' Principle tells us that the propagator takes a particularly simple form, and we can write

$$E_\nu(\mathbf{r}) = \int \mathcal{E}_\nu(\mathbf{R}) \frac{e^{2\pi i \nu |\mathbf{R}-\mathbf{r}|/c}}{|\mathbf{R}-\mathbf{r}|} \, dS \, . \tag{1–2}$$

Here dS is the element of surface area on the celestial sphere.

Equation 1–2 is the general form of the quasi-monochromatic component of the electric field at frequency ν due to all sources of cosmic electromagnetic

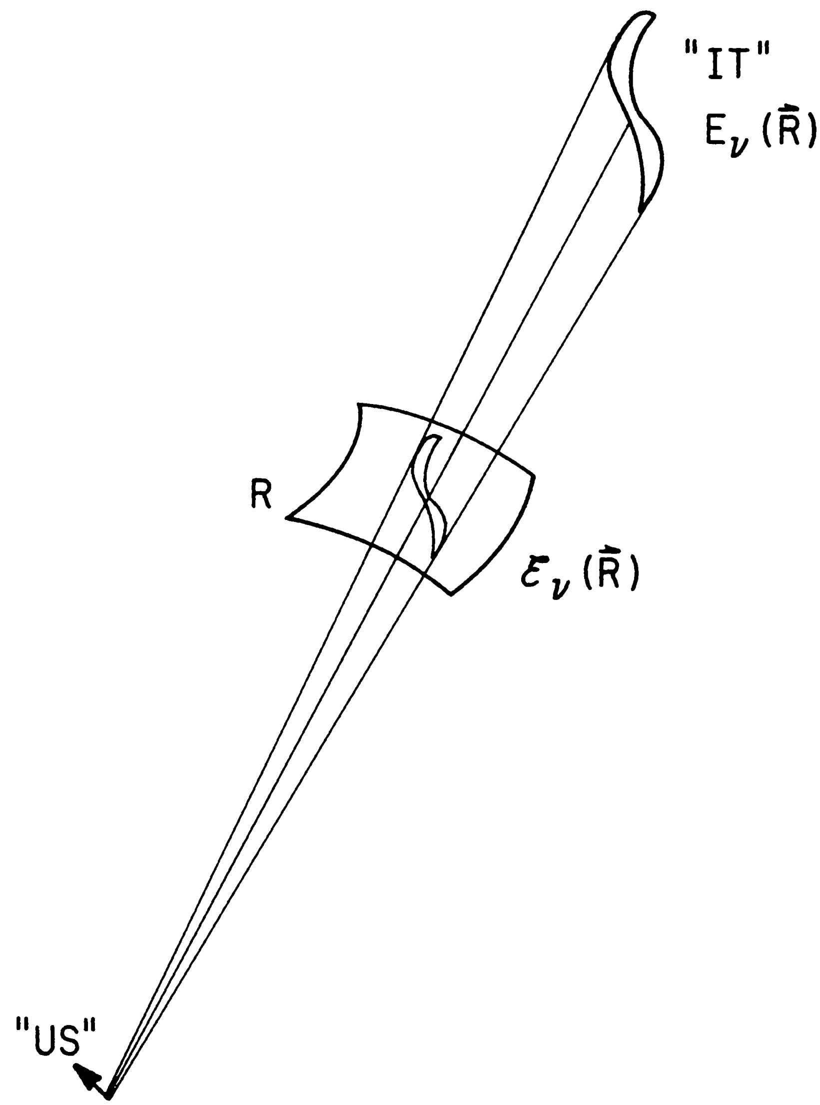

Figure 1–2. The passage of radiation from a distant source through an imaginary sphere at radius $|\mathbf{R}|$ defines the electric field distribution $\mathcal{E}_\nu(\mathbf{R})$ on this surface. For the purposes of synthesis imaging, all astronomical sources may be considered to lie on this sphere, provided only that their real distances greatly exceed B^2/λ, where $B = |\mathbf{r_1} - \mathbf{r_2}|$ is the baseline length.

radiation. This is all we have; we can measure only the properties of this field $E_\nu(\mathbf{r})$ to tell us about the nature of things at large in the universe.

3. Spatial Coherence Function of the Field

Among the properties of $E_\nu(\mathbf{r})$ 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 a product, namely $\mathsf{V}_\nu(\mathbf{r_1}, \mathbf{r_2}) = \langle \mathbf{E}_\nu(\mathbf{r_1})\mathbf{E}^*_\nu(\mathbf{r_2})\rangle$. The raised asterisk indicates the complex conjugate. We can then use Equation 1–2 to substitute for $E_\nu(\mathbf{r})$, writing the product of the integrals as a double integral over two separate surface element dummy variables:

$$\mathsf{V}_\nu(\mathbf{r}_1, \mathbf{r}_2) = \left\langle \iint \mathcal{E}_\nu(\mathbf{R}_1)\mathcal{E}_\nu^*(\mathbf{R}_2) \frac{e^{2\pi i\nu|\mathbf{R}_1 - \mathbf{r}_1|/c}}{|\mathbf{R}_1 - \mathbf{r}_1|} \frac{e^{-2\pi i\nu|\mathbf{R}_2 - \mathbf{r}_2|/c}}{|\mathbf{R}_2 - \mathbf{r}_2|} \, dS_1 \, dS_2 \right\rangle . \tag{1–3}$$

The fourth simplifying assumption is that the radiation from astronomical objects is not spatially coherent; i.e., that $\langle \mathcal{E}_\nu(\mathbf{R}_1)\mathcal{E}_\nu^*(\mathbf{R}_2)\rangle$ is zero for $\mathbf{R}_1 \neq \mathbf{R}_2$. Exchanging the expectation and the integrals in Equation 1–3 then gives:

$$\mathsf{V}_\nu(\mathbf{r}_1, \mathbf{r}_2) = \int \langle|\mathcal{E}_\nu(\mathbf{R})|^2\rangle |\mathbf{R}|^2 \frac{e^{2\pi i\nu|\mathbf{R} - \mathbf{r}_1|/c}}{|\mathbf{R} - \mathbf{r}_1|} \frac{e^{-2\pi i\nu|\mathbf{R} - \mathbf{r}_2|/c}}{|\mathbf{R} - \mathbf{r}_2|} \, dS \, . \tag{1–4}$$

Now write $\mathbf{s}$ for the unit vector $\mathbf{R}/|\mathbf{R}|$ and $I_\nu(\mathbf{s})$ for the observed *intensity* $|\mathbf{R}|^2\langle|\mathcal{E}_\nu(\mathbf{s})|^2\rangle$. The second assumption (the great distance to the sources and to the celestial sphere) can then be used again to neglect the small terms of order $|\mathbf{r}/\mathbf{R}|$, and to replace the surface element dS on the celestial sphere by $|\mathbf{R}|^2 d\Omega$, so that Equation 1–4 becomes:

$$\mathsf{V}_\nu(\mathbf{r}_1, \mathbf{r}_2) \approx \int I_\nu(\mathbf{s}) e^{-2\pi i\nu \mathbf{s}\cdot(\mathbf{r}_1 - \mathbf{r}_2)/c} \, d\Omega \, . \tag{1–5}$$

Observe that Equation 1–5 depends only on the separation vector $\mathbf{r}_1 - \mathbf{r}_2$ of the two points, not on their 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 a single (vector) separation $\mathbf{r}_1 - \mathbf{r}_2$ is called the *spatial coherence function*, or the *spatial autocorrelation function*, of the field $E_\nu(\mathbf{r})$. It is all we have to measure.

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

4. The Basic Fourier Inversions of Synthesis Imaging

A second interesting point about Equation 1–5 is that the equation 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.

There are two special cases of a great deal of interest. In fact, it is usually argued that any actual case is so close to one of these two special cases that the invertibility properties (although not necessarily the effort required to perform the inversion) must be essentially similar. Since there are two forms of interest, there are two alternate forms of our fifth (and final) simplifying assumption.

4.1. Measurements confined to a plane

First, we could choose to make our measurements only in a plane; that is, in some favored coordinate system, the vector spacing of the separation variable in the coherence function, conveniently measured in terms of the wavelength $\lambda = c/\nu$, is $\mathbf{r}_1 - \mathbf{r}_2 = \lambda(u, v, 0)$. In this same coordinate system, the components of the unit vector $\mathbf{s}$ are $(l, m, \sqrt{1 - l^2 - m^2})$. Inserting these, and explicitly showing the form, in this coordinate system, of the element of solid angle, we have

$$V_\nu(u, v, w \equiv 0) = \iint I_\nu(l, m) \frac{e^{-2\pi i(ul+vm)}}{\sqrt{1 - l^2 - m^2}}\, dl\, dm\,. \tag{1–6}$$

Equation 1–6 is, clearly, a Fourier transform relation between the spatial coherence function $V_\nu(u, v, w \equiv 0)$ (with separations expressed in wavelengths), and the modified intensity $I_\nu(l, m)/\sqrt{1 - l^2 - m^2}$ (with angles expressed as direction cosines). Now we are home free. Mathematicians have devoted decades of work to telling us when we can invert a Fourier transform, and how much information it requires.

4.2. All sources in a small region of sky

The alternate form of the fifth simplifying assumption is to consider the case where all of the radiation comes from only a small portion of the celestial sphere. That is, to take $\mathbf{s} = \mathbf{s}_0 + \sigma$, and neglect all terms in the squares of the components of σ. In particular, the statement that both $\mathbf{s}$ and $\mathbf{s}_0$ are unit vectors implies that

$$\begin{aligned} 1 = |\mathbf{s}| = \mathbf{s} \cdot \mathbf{s} &= \mathbf{s}_0 \cdot \mathbf{s}_0 \\ &= \mathbf{s}_0 \cdot \mathbf{s}_0 + 2\mathbf{s}_0 \cdot \sigma + \sigma \cdot \sigma \\ &\approx 1 + 2\mathbf{s}_0 \cdot \sigma\,, \end{aligned}$$

i.e., $\mathbf{s}_0$ and σ are perpendicular. If we again introduce a special coordinate system such that $\mathbf{s}_0 = (0, 0, 1)$, then we have a slightly different offspring of Equation 1–5:

$$V'_\nu(u, v, w) = e^{-2\pi i w} \iint I_\nu(l, m) e^{-2\pi i(ul+vm)}\, dl\, dm\,. \tag{1–7}$$

Here, the components of the vector $\mathbf{r}_1 - \mathbf{r}_2$ have been denoted by (u, v, w). It is customary to absorb the factor in front of the integral in Equation 1–7 into the left hand side, by considering the quantity $V_\nu(u, v, w) = e^{2\pi i w} V'_\nu(u, v, w)$, which we see is independent of w:

$$V_\nu(u, v) = \iint I_\nu(l, m) e^{-2\pi i(ul+vm)}\, dl\, dm\,. \tag{1–8}$$

$V_\nu(u, v)$ 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, we have in particular, the direct inversion

$$I_\nu(l, m) = \iint V_\nu(u, v) e^{2\pi i(ul+vm)}\, du\, dv\,. \tag{1–9}$$

The relationship between the two different forms of the assumption used here and in Section 4.1 can be seen from the symmetric role played in Equation 1–5 by the two vectors $\mathbf{s}$ and $\mathbf{r}_1 - \mathbf{r}_2$: the form used in Section 4.1 results from saying that the *vectors* $\mathbf{r}_1 - \mathbf{r}_2$ lie in a plane; the form used here results from saying that the *endpoints* of the vectors $\mathbf{s}$ lie in a plane.

4.3. Effect of discrete sampling

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. One can then calculate a function

$$I_\nu^D(l,m) = \iint V_\nu(u,v)S(u,v)e^{2\pi i(ul+vm)}\,du\,dv\,. \qquad (1\text{–}10)$$

Radio astronomers often refer to $I_\nu^D(l,m)$ as the *dirty image*; its relation to the desired intensity distribution $I_\nu(l,m)$ is (using the convolution theorem for Fourier transforms):

$$I_\nu^D = I_\nu * B\,, \qquad (1\text{–}11)$$

where the in-line asterisk denotes convolution and

$$B(l,m) = \iint S(u,v)e^{2\pi i(ul+vm)}\,du\,dv \qquad (1\text{–}12)$$

is the *synthesized beam* or *point spread function* corresponding to the sampling function $S(u,v)$. Equation 1–11 says that I^D is the true intensity distribution I convolved with the synthesized beam B. Lecture 8 discusses methods for undoing this convolution.

4.4. Effect of the element reception pattern

An additional minor alteration must be made to the above for convenience in practical calculation. In practice, the interferometer elements are not point probes which sense the voltage at that point, but are elements of finite size, which have some sensitivity to the direction of arrival of the radio radiation. That is, there is an additional factor within the integral of Equation 1–2, and hence of Equations 1–4, 1–5, 1–6, 1–7 and 1–8, of $\mathcal{A}_\nu(\mathbf{s})$ (the *primary beam* or *normalized reception pattern* of the interferometer elements) describing this sensitivity as a function of direction. For explicitness, Equation 1–8 is rewritten below, with this factor included:

$$V_\nu(u,v) = \iint \mathcal{A}_\nu(l,m)I_\nu(l,m)e^{-2\pi i(ul+vm)}\,dl\,dm\,. \qquad (1\text{–}13)$$

The $V_\nu(u,v)$ so defined is normally termed the *complex visibility* relative to the chosen phase tracking center.

It is clear that dealing with the element directivity $\mathcal{A}_\nu$ should be postponed to the final step of deriving the sky intensity, and that then it should simply divide the derived intensities (if all elements have the same reception pattern). This division will, however, not only produce a better estimate of the actual intensities in this direction, but will also increase the errors (of all types) in directions far from the center of the element primary beam, where one is dividing by small numbers.

Although the factor $\mathcal{A}_\nu$ looks like merely a nuisance, it is actually the reason that the second form of the final assumption (used in Section 4.2) is so acceptable in many practical cases—$\mathcal{A}_\nu(\mathbf{s})$ falls rapidly to zero except in the vicinity of some $\mathbf{s}_0$, the pointing center for the array elements.

5. Extensions to the Basic Theory

Two simple extensions to this basic theory are worth mentioning at this point.

5.1. Spectroscopy

First, consider the case of observing a spectral line. Here the appearance of the sky may change quite rapidly as a function of frequency, and one would like to make synthesis images at a large number of closely spaced frequencies. Clearly, one can do this by inserting narrowband filters into the early, linear, parts of the interferometer, and simply repeat the processing for each frequency, either seriatim or simultaneously. However, there is a technically more attractive approach. With current technology, it is attractive to implement the latter portions of the interferometer in digital hardware. In this technology, it is quite inexpensive to add additional multipliers to calculate the correlation as a function of lag. Admitting a range of quasi-monochromatic waves to the interferometer, we can write an expression for the correlation as a function of lag, noting that for each quasi-monochromatic wave, a lag is equivalent to a phase shift, i.e., a multiplication by a complex exponential

$$V(u,v,\tau) = \int V(u,v,\nu) e^{2\pi i \tau \nu} \, d\nu \, . \tag{1–14}$$

The above is clearly a Fourier transform, with complementary variables ν and τ, and can be inverted to extract the desired $V(u,v,\nu)$. Since, in this digital technology, one is dealing with sampled data, I give the sampled form of the inversion below:

$$V(u,v,j\Delta\nu) = \sum_k V(u,v,k\Delta\tau) e^{-2\pi i jk \Delta\nu \Delta\tau} \, . \tag{1–15}$$

The fact that we are dealing with sampled data is of some interest, and we should stop and inquire about how the Fourier sampling theorem is to be applied. Examining the above, in its full complex form, we see that the replication interval is $1/\Delta\tau$ in frequency, so that the band of frequencies must be limited, before sampling, to a total bandwidth of less than this, to avoid loss of information in the sampling process.

This is different from the statement one usually encounters, in which a prefiltering to $1/2\Delta\tau$ is required to preserve the information in the sampling process for a signal (actually it is usually stated, equivalently, as requiring a sampling interval of $1/2B$, where B is the prefilter bandwidth). This factor of two difference is due to the complex nature of the quantities we have been dealing with—the $V(u,v,\nu)$ are complex numbers, calculated by a complex multiplication of the complex field quantities. Complex multipliers and complex samplers require at least twice as many electronic components as devices that produce a real number, and the resulting doubling of the hardware permits us to sample a factor of two less often.

However, one can also develop this theory from the conventional viewpoint of dealing with real numbers only. Here the $2B$ sample rate is required, and maintains all the information in the signals. We can exchange this faster sampling rate for the double hardware needed to produce the complex version of the signals. The real parts of the various $V(u,v,\nu)$ are derived from the part of the correlation function that is even about $\tau = 0$, and the odd part supplies the imaginary part of $V(u,v,\nu)$.

Finally, if one derives the spectrum in this manner, one can, clearly, convert back to the single continuum channel at zero lag simply by summing the derived frequency-dependent V. This process clearly results in a complex number, even though each measurement was only a real number. The process of transforming a real function into a complex one by Fourier transforming and then transforming back on half the interval is called a Hilbert transform, and is an alternate method to implementing complex correlators.

5.2. Polarimetry

Now, in a final remark, let me look back to Section 2, to the first simplifying assumption, that of a scalar field. Actually, the electromagnetic field is a vector phenomenon, and the polarization properties carry interesting physical information. For the case of noise emission, one must be a bit careful about the definition of polarization. A monochromatic wave is always completely polarized, in some particular elliptical polarization, in that a single number describes the variation of the fields everywhere. For electromagnetic noise, polarization is defined by a correlation process. One picks two orthogonal polarizations and analyzes the radiation of the quasi-monochromatic waves into the components in these two polarizations. Then 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, then the matrix

$$\begin{pmatrix} RR^* & RL^* \\ LR^* & LL^* \end{pmatrix} \tag{1–16}$$

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

$$\begin{pmatrix} I+V & Q+iU \\ Q-iU & I-V \end{pmatrix} \tag{1–17}$$

The complex correlation functions on the celestial sphere are preserved in the spatial coherence functions that interferometers measure. That is, one can derive, for instance, the distribution of $\langle RR^* \rangle$ on the sky by measuring the coherence function of R on the ground, and so forth for the other components of the matrix in (1–16). Since the intensity is the quantity in which one is always interested, one usually forms the sum of the R and L coherence functions before transforming to the sky plane, which one can always do, since the relations are linear. One can choose to do the same with the other Stokes parameters, or one can calculate the transforms of the mutual coherence between R and L to find the distribution of $\langle RL^* \rangle$ on the sky, and later note that this is, in terms of the Stokes parameters, $Q + iU$.

References

Born, M. & Wolf, E. 1980, *Principles of Optics*, Sixth (Corrected) Edition, Pergamon Press (Oxford, England).

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

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

Swenson, G. W., Jr. & Mathur, N. C. 1968, *Proc. IEEE*, 56, 2114–2130.

Yen, J. L. 1985, in *Array Signal Processing*, S. Haykin, Ed., Prentice–Hall (Englewood Cliffs, NJ), pp. 293–350.