PART I

APERTURE SYNTHESIS METHODS

https://doi.org/10.1017/S0252921100074686 Published online by Cambridge University Press

FUNDAMENTALS AND DEFICIENCIES OF APERTURE SYNTHESIS

(Invited paper)

Edward B. Fomalont
National Radio Astronomy Observatory
Green Bank, West Virginia 24944 U.S.A.

1. INTRODUCTION

Aperture synthesis is the method used by astronomers to determine the accurate brightness distribution of the radio sky with a resolution much better than that possible with a single large antenna. The technique, now over a decade old, utilizes a large number of connected radio antennas, some of them physically moveable, to follow a region of sky for many hours or days in order to sample the spatial coherence function of the radiation field over a sufficiently large area and with a reasonable filling factor. Landmark references for aperture synthesis are McCready et al. (1947), Stanier (1950), Christiansen and Warburton (1955), Lequeux et al. (1962), Read (1961) and Ryle and Hewish (1960).

In Section 2 the relationship between the spatial coherence function and the brightness distribution, essentially the van Cittert-Zernike Theorem, together with necessary assumptions about the radiation field, will be discussed. Aperture synthesis is a straightforward application of this theorem. The deficiencies in aperture synthesis are many and these are summarized in Section 3. A major part of this colloquium will deal with methods to correct the adverse effects of these deficiencies. Some are fundamental to the technique--e.g., the spotty sampling of the spatial coherence function. Some are related to practical considerations in the design of the array and the speed of map making at the expense of some accuracy--e.g., the use of the Fast-Fourier transform algorithm. Some are related to non-stationary effects during the course of the observations--e.g., tropospheric phase fluctuations. In Section 4 a brief description of problems in aperture synthesis that strike the author as important or annoying as well as a *potpourri* of other ideas will be mentioned.

C. van Schooneveld (ed.), Image Formation from Coherence Functions in Astronomy, 3-18.

2. FUNDAMENTALS OF APERTURE SYNTHESIS

2.1. Coherence Properties of the Radiation Field

Let $E(\bar{u},t)$ represent the radiation field at any point in space, $\bar{u}$, at any time, t. Although electromagnetic radiation is described by a vector field which satisfies Maxwell's Equations, for most astronomical applications an approximate description of the field by a scalar wave function (for example, the transverse component of the electric field) is adequate. The field fluctuates rapidly and behaves ergodically; that is, the time-average properties are well-defined and any measurement of an average property is typical of all such similar measurements.

While it is now possible at radio frequencies to follow the detailed fluctuations of the field, there is little additional information in them and most characteristics of the radiation are embodied in various average functions of the field. The intensity of the field can be easily measured using a suitable probe which responds with a signal $S(\bar{u},t)$ proportional to $E(\bar{u},t)$. The signal is first "detected" by measuring the power and then averaged over some time interval T which is long compared with the time-scale of the fluctuations. The intensity $I(\bar{u})$ is thus

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} S^*(\bar{u},t)\, S(\bar{u},t)\, dt \tag{2.1}$$

and is usually written in short-hand notation as

$$I(\bar{u}) = \langle S^*(\bar{u},t)\, S(\bar{u},t)\rangle \ . \tag{2.2}$$

As is customary when dealing with signals associated with the wave equation, a complex number, called the analytic signal, is used in place of the real signal (e.g., Steel 1967, p.22).

When two or more sources of radiation are superimposed the intensity in the region of superposition can vary. This phenomenon is called interference. In general the radiations from different sources are incoherent; i.e. their fluctuations are completely uncorrelated and, thus, they do not interfere over typical averaging time scales $t > 1/\Delta\nu$ where $\Delta\nu$ is the bandwidth of the radiation. But if the radiation from *one* source is separated into two beams, by sampling two points of the wavefront or by splitting the radiation at some point, the fluctuations in the two beams are generally correlated and they will interfere when combined.

The measure of the correlation in a radiation field between two points $\bar{u}_1$ and $\bar{u}_2$ at a time difference τ is given by

$$\Gamma(\bar{u}_1,\bar{u}_2,\tau) = \langle S^*(\bar{u}_1,t) \quad S(\bar{u}_2,\ t+\tau)\rangle \tag{2.3}$$

and is denoted as the *mutual coherence function* of the field. For example if we combine the signals sampled at two points $\bar{u}_1$ and $\bar{u}_2$, separated by time τ, the time-average intensity of the signal $S_s = S(\bar{u}_1,t) + S(\bar{u}_2,t)$ would be

$$\langle S_s^* \ S_s\rangle = I(\bar{u}_1) + I(\bar{u}_2) + 2 \text{ Real part } \Gamma(\bar{u}_1,\bar{u}_2,\tau). \tag{2.4}$$

The measured intensity could have any value between 0 and 4I depending on the phase and amplitude of the mutual coherence function. The last term in equation (2.4) can be isolated by multiplying, rather than adding, the two signals; the imaginary part can be obtained by introducing a 90-deg phase shift in one of the signals. Thus the mutual coherence function of a radiation field can be easily measured.

2.2. The Van Cittert-Zernike Theorem

The arrays used in radio astronomy sample many parts of the wavefront of the radiation from a source. Thus, the sum (or product) of signals from any pair of antennas is related to $\Gamma(\bar{u}_1,\bar{u}_2,\tau)$ where $\bar{u}_1$ is the position of one antenna, $\bar{u}_2$ is the position of the other antenna and τ is the time difference between the samples measured with respect to a wavefront from some arbitrary direction. If the two points coincide, $\Gamma(\bar{u},\bar{u},\tau)$ is called the *autocorrelation function* of the field at $\bar{u}$. By the Wiener-Khinchin Theorem the autocorrelation function is the Fourier transform of the power spectrum of the field (e.g., Steel 1967, p. 44). On the other hand for the special case when τ=0, $\Gamma(\bar{u}_1,\bar{u}_2,0)$ is called the *spatial coherence function* of the field.

The relationship between the spatial coherence function and the brightness distribution of an extended source is given by the van Cittert-Zernike Theorem. A derivation is given by Born and Wolf (1964, p. 510) or Steel (1967, p. 47). The theorem applies only if the radiation from the source is incoherent; that is the fluctuations of the radiation from different elements of the source are statistically independent.

For most astronomical applications, a radio source can be considered in the far field and the van Cittert-Zernike Theorem reduces to

$$\Gamma(\bar{u},\tau) = \int_{\nu=0}^{\infty} \left\{ \exp(i2\pi\nu\tau) \int_{\bar{\ell}} I(\bar{\ell},\nu)\exp(i2\pi\frac{\nu}{c}\ \bar{u}\cdot\bar{\ell})d\bar{\ell} \right\} d\nu \tag{2.5}$$

https://doi.org/10.1017/S0252921100074686 Published online by Cambridge University Press

where $\bar{u}$ = the separation of the sampled field points = $\bar{u}_1-\bar{u}_2$, independent of origin
$\bar{\ell}$ = the direction cosine to an element of the source
ν = the frequency of the radiation
c = the speed of light
$I(\bar{\ell},\nu)$ = the brightness distribution of the source
$\Gamma(\bar{u},\tau)$ = the mutual coherence function.

The actual measurement of the spatial coherence function $\Gamma(\bar{u},0)$ in typical radio astronomy applications is complicated by the following: 1) the radiation probes (antennas) sample the field over a finite area with relative sensitivity $a(\bar{u})$, usually called the grading of the antenna; 2) the bandwidth of radiation accepted is finite with a relative sensitivity given by $B(\nu)$; 3) all signals must be combined with delay $\tau=0$ with respect to some arbitrary direction, usually denoted as the phase center or map center. With these practical considerations, the measured response of the correlation of the signals separated by $\bar{u}$ is called the *visibility function* and is given by

$$V(u,v,w) = \int_{\nu=0}^{\infty} B(\nu) \left\{ \int_{\ell,m} \frac{1}{n} A(\ell,m,\nu)\ I(\ell,m,\nu) \exp\left\{ i2\pi\frac{\nu}{c}\ [u\ell+vm+w(1-n)]\right\}\ d\ell dm \right\} d\nu \tag{2.6}$$

We have resolved $\bar{\ell}$, the direction cosine to the source with respect to the phase center and $\bar{u}$, the separation of the two samples of radiation into the usual components

$\bar{\ell}$ = (ℓ,m,n) where the n-axis is parallel to the phase center direction, the ℓ- and m-axes define the plane parallel to n, ℓ directed to the east and m directed to the north. [Because the radio source is confined to the unit sphere, $n^2 = 1-\ell^2-m^2$.]

$\bar{u}$ = (u,v,w) where $u\|\ell$, $v\|m$, and $w\|n$.

The effect of the antenna grading is the multiplicative factor $A(\ell,m,\nu)$, called the antenna field pattern. If conventional radio telescopes are used as the radiation probe, A is near zero beyond a radius ℓ_{max} so that little information about $I(\ell,m,\nu)$ is measured beyond a *field of view* of radius ℓ_{max}.

2.3. Solution for the Brightness Distribution

It is clear from equation (2.6) that the visibility function measures something like a Fourier component of the brightness distribution. Aperture synthesis is, essentially, a methodical technique whereby a large number of samples in (u,v,w) are measured by 1) correlating the signals from all pairs (or at least all non-redundant pairs) of antennas in the multi-element array, 2) physically moving antennas to change the spacings between pairs of antennas, and 3) utilizing the rotation of the Earth to change the source-array aspect. With sufficient sampling, it is then possible to estimate the brightness distribution from the measured set of visibility functions.

However, an explicit expression for the brightness distribution can only be obtained with two further simplifications. In equation (2.6) the frequency ν and the angular coordinate $\bar{\ell}$ are coupled in the exponential term and cannot in general be separated. A quasi-monochromatic approximation applies when the change in phase of the exponential term is negligible over the radiation bandwidth $\Delta\nu$, which gives the condition

$$\frac{\Delta\nu}{c} u_{max} \, \ell_{max} << 1 \tag{2.7}$$

where u_{max} = maximum baseline of the array
ℓ_{max} = maximum angular extent of the emission, usually limited by the field pattern of the antenna.

We may then define an *apparent brightness distribution* $I'(\ell,m)$ given by

$$I'(\ell,m) = \int_{\nu} B(\nu) \, \frac{1}{n} \, A(\ell,m,\nu) \; I(\ell,m,\nu) \; d\nu. \tag{2.8}$$

The apparent brightness distribution is now a function of the array parameters as well as the radiation from the sky. In order to make any Fourier inversion of equation (2.6) meaningful the apparent brightness distribution must be invariant for all antenna pairs and be unchanged for the duration of the observations. This is accomplished, in part, by building an array with identical antennas and supporting electronics. Stringent mechanical and electrical tolerances are also necessary to achieve the necessary stability of I' to permit accurate aperture synthesis. Particular problems will be discussed in Section 4.

The explicit expression of the apparent brightness distribution in terms of the visibility function can now be written

$$I'(\ell,m) \; \delta(n-n') = \int V(u,v,w) \exp\left\{-i2\pi\frac{\nu_0}{c}\,(u\ell+vm+w[1-n])\right\} dudvdw \tag{2.9}$$

where $n' = (1-l^2-m^2)^{\frac{1}{2}}$ so that the emission is confined to the unit sphere, δ is the Dirac-delta function and ν_0 is the average frequency of I'.

https://doi.org/10.1017/S0252921100074686 Published online by Cambridge University Press

The assumptions made in obtaining equation (2.9) are:

(1) The radiation is ergodic
(2) The radiation is incoherent across the source
(3) The radio source is in the far field
(4) The signals of each pair are correlated with $\tau=0$
(5) The quasi-monochromatic approximation applies
(6) $I'(\ell,m)$ is well-defined and stationary.

The breakdown of assumptions 1 and 2 would be surprising except for non-natural signals. The limitation imposed by the far field assumption can be lessened if we know the distance to the radio source and can correct for the effects of the spherical wavefront in the measured visibility. Many of the problems of high resolution, high frequency aperture synthesis arise from a breakdown of assumption 4 because of the variable differential delay across an array caused by the propagation of radiation in the troposphere of the Earth. If assumption 5 is not valid, then there is a loss of information about the brightness distribution at large angular separations from the phase center. The degree of validity of assumption 6 is a large measure of the ultimate accuracy in which the brightness distribution can be recovered from the visibility function.

The apparent brightness distribution and the visibility function are three-dimensional Fourier pairs; however, since I' is only defined on the unit sphere, it is often possible to reduce the dimensionality of the inversion formula to two. For many aperture synthesis applications the term $w(1-n)\ \nu_0/c \ll 1$ anywhere in the field of view and can be neglected. The three dimensional inversion can also be avoided if the sampled visibility function is coplanar over the entire set of observations. In this case $w=au+bv$ where a and b are arbitrary constants and a redefinition of ℓ and m gives the relation

$$I'(\ell',m') = \int V(u,v,w)\exp\left\{-i2\pi\frac{\nu_0}{c}\ (u\ell'+vm')\right\}\ dudv \tag{2.10}$$

with ℓ' and m' now not quite direction cosines with respect to the phase center. Brouw (1971) was the first to show that only an east-west array (or an array on a plane perpendicular to the rotation axis of the earth) describes a coplanar aperture under earth-rotation synthesis. Other ground-based arrays may, at any instant, sample the visibility function on a plane and may use a two-dimensional transform for data taken over a short duration. However, the combination of maps over many hours of earth-rotation synthesis require a continual redefinition of ℓ and m. For ground-based arrays which are not east-west and larger than several kilometers, the curvature of the earth invalidates the coplanar property of the array.

The measurement of the polarization characteristics of the radiation can be obtained by using probes which are sensitive to the various components of the field. No additional problems are encountered in aperture synthesis as long as the entire set of observations are consistent in their polarization properties.

The mutual coherence function may be described as a second-order correlation function of the field since it involves two signals. A generalization to higher-order correlators is possible (Glauber 1963). An example is the intensity interferometer (Brown and Twiss 1954). The instantaneous signal at two points is *first* detected and *then* multiplied together to give the correlation. The response R is

$$R(\bar{u}_1,\bar{u}_2,\tau) = \langle S^*(\bar{u}_1,t)\ S(\bar{u}_1,t)\ S^*(\bar{u}_2,t+\tau)\ S(\bar{u}_2,t+\tau)\rangle \tag{2.11}$$

and can be shown to be equal to (Mandel 1963)

$$R(\bar{u}_1,\bar{u}_2,\tau) = I(\bar{u}_1)I(\bar{u}_2) + \left|\Gamma(\bar{u}_1,\bar{u}_2,\tau)\right|^2. \tag{2.12}$$

Thus, only the amplitude of the mutual coherence function is obtained.

3. DEFICIENCIES OF APERTURE SYNTHESIS

3.1. Undersampling and the Two Basic Reconstruction Methods

Perhaps the most annoying and fundamental deficiency in aperture synthesis is the undersampling of the visibility function in the (u,v) plane. (For simplicity equation (2.10), the two-dimensional inversion formula will be used as the explicit solution for I', realizing that equation (2.9) should be used if necessary.) The determination of I' explicitly demands the value of the visibility function measured continuously within some region. Instead, it is sampled at discrete points within a radius of $\sim u_{max}$, along an elliptical track (one for each pair of antennas) in the (u,v) plane as the source moves in a diurnal path, with noticable omissions in coverage such as missing short spacings, a missing angular wedge, and other coverage anomalies such as perfectly-scaled tracks or regions of particularly dense sampling.

The conventional method of determining I' is the direct Fourier Transform (DFT) in which equation (2.10) is summed only for the P samples of the visibility function.

$$I''(\ell,m) = \sum_p \alpha(u_p,v_p)\ V(u_p,v_p)\exp\left\{-2\pi\frac{\nu_0}{c}\ (u_p\ell+v_pm)\right\}\quad 1<p<P. \tag{3.1}$$

The *reconstructed image* I" is not in general a particularly good representation of I' but it does have the property that

https://doi.org/10.1017/S0252921100074686 Published online by Cambridge University Press

$$I''(\ell,m) = I'(\ell,m) * B(\ell,m) \tag{3.2}$$

with

$$B(\ell,m) = \sum_p \alpha(u_p,v_p)\exp\left\{-i2\pi\frac{\nu_0}{c}\,(u_p\ell+v_pm)\right\} \quad 1<p<P \tag{3.3}$$

where $B(\ell,m)$ is the response to a point source (point-spread function or dirty beam) and $*$ means convolution. The apodizing function or weights $\alpha(u_p,v_p)$ can be chosen arbitrarily and are used to change some of the characteristics of I" within certain limits. For arrays composed of a line of antennas equation (3.1) can be conveniently summed in the radial and then azimuthal coordinates in the (u,v) plane (Kenderdine 1974). This is equivalent to obtaining a one-dimensional strip brightness distribution of the radio source and then summing the strip distributions over angle. Algebraic reconstruction techniques for this problem have been studied in electron microscopy (e.g., Gordon 1974).

The restoration of I' from I" is a problem of deconvolution (Högbom 1974). The image or map I" and the beam B contain ripples, called *sidelobes*, produced by spotty sampling of the visibility function. The sidelobes come in all shapes and sizes and the major features can be traced to particularly large gaps or noticable periodicities in the (u,v) coverage.

Non-Fourier techniques utilize equation (2.6), with I' defined by equation (2.8), to determine an I' which reproduces the observed visibility function at the sampled points. For example, model-fitting techniques (Fomalont and Wright 1974) start with the assumption that I' can be approximated by a small number of discrete components, usually Gaussian-shaped. The parameters which define the strength, position and size of the components are then adjusted using a search technique, (the simplex method has been found useful in this regard) to better fit the observed data (e.g., Nelder and Mead 1965). In the maximum entropy method (MEM; Ables 1974) a solution of I', usually represented in a pixel form, is obtained which satisfies equation (2.6) and maximizes the entropy of I'. The use of non-Fourier techniques, especially MEM, is becoming more prevalent as more efficient algorithms are being developed and as the properties of these methods are better understood. In addition, certain classes of errors associated with the measurement of the visibility function or with the addition of *a priori* knowledge of the nature of I' cannot be incorporated well in the standard Fourier methods.

https://doi.org/10.1017/S0252921100074686 Published online by Cambridge University Press

3.2. Wide Bandwidths

The basic equations of aperture synthesis were derived for the quasi-monochromatic case of relatively narrow bandwidths. However, this constraint is not satisfied for many large arrays which use wide bandwidths to increase sensitivity. In general there is no way of recovering I' from V [using measurements of the spatial coherence function] without unrecoverable loss of resolution. The mutual coherence function (visibility function at many delay lags) must be sampled from which $I'(\ell,m,\nu)$ can be determined in narrow frequency bands, each of which satisfy the quasi-monochromatic assumption. The inversion formulae would be similar to equation (2.9) with the addition of another dimension of time t with the (u,v,w) domain and frequency ν with the (ℓ,m,n) domain.

Application of equation (3.1) in the wide bandwidth case leads to a reconstructed image which is the result of summing the apparent brightness distribution at each frequency but with the angular scale multiplied by ν/ν_0. This produces a radial smearing in I" apart from the other beam convolution effects. The extent of the radial smearing is proportional to the distance of the source from the phase center and the bandwidth. However, I" is still a convolution of the point source response

$$I'' = I'(\text{radially smeared}) * B \tag{3.4}$$

and standard deconvolution techniques are still useful. In a sense, no information is lost when wide bandwidths are observed; only the radial resolution degrades at angular separations far from the phase center.

3.3. Practical Considerations in Fourier Inversions

The DFT is extremely time consuming to calculate in a digital computer. Significant computation labor is saved by use of the Fast-Fourier Transform algorithm (FFT) which requires the visibility function to be sampled on an equi-spaced grid in the (u,v) plane (Cochran et al. 1967). The transformation of the originally sampled visibility function (by a convolution in the (u,v) plane and then a further sampling at the appropriate grid points) further distorts the reconstructed map (e.g., Fomalont 1973) as compared with that obtained with the DFT algorithm. Perhaps, the most annoying effect of the FFT is the *aliasing* caused by the periodic sampling. If I' contains significant response outside of the area in the field of view of the FFT (either real emission or sidelobes of I') this response will be reflected back into the FFT field of view with an amplitude related to the convolution function. Also, the errors in the reconstructed map will not be as uniformly distributed as with the DFT.

It is possible to use an optical system to perform the Fourier transform and a possible scheme has been investigated (ERIM report to VLA 1977). The visibility function is encoded on a piece of film, coherent light is then passed through the film and focussed by a lens onto an array of sensors. The method works but the accuracy of inversion is far surpassed by digital techniques.

3.4. Propagation Effects

A major source of error in aperture synthesis (at least ground-based aperture synthesis) is produced by changes in the propagation properties of the troposphere and ionosphere (Basart, Miley and Clark 1970, Hamaker 1978). For baselines longer than ∿ 1 km the radiation wavefront is continually distorted over minutes and hours by refraction from the dry-air component and the water-vapor component of the troposphere. At frequencies above 2 GHz the resulting differential delay of the radiation between the antenna pairs causes large phase errors in the measured visibility function. At frequencies above about 10 GHz and for baselines longer than 20 km, the phase of the visibility function may be virtually useless in some weather conditions. Generally, the visibility amplitude is unaffected except at frequencies above 5 GHz where water vapor absorption becomes important. At frequencies below about 0.3 GHz, the ionosphere of the earth produces similar fluctuations in the measured phase.

With increasing emphasis on high frequency, high resolution aperture synthesis reconstruction and restoration techniques which can handle data with large phase errors are necessary. Recent advances and understandings of various techniques will be reported during this colloquium.

3.5. Errors and Noise

All of the observations contain random short term errors caused by receiver noise and radiation from the sky and ground. These errors have the usual statistical behavior and their effects are easily analyzed in terms of the sensitivity limits of the observations.

Other errors produce variations of I' (see assumption 6 in Section 2.3) between the various antenna pairs and temporal changes caused by various instrumental imperfections. Obviously the design considerations for an array attempt to balance a loss of accuracy in determining I with various technical and economical considerations.

As a general rule, errors in the measured visibility function which vary slowly in time cause more grief than somewhat larger errors which have shorter characteristic time scales. For example, a pointing error in an antenna of an amount X of short term duration caused by gusty wind conditions would not have as large an effect on the accuracy of a radio map as a gradually changing pointing error much less than X lasting over the duration of the observations. It is often difficult to estimate the limitation in the brightness distribution accuracy caused by these non-random errors.

https://doi.org/10.1017/S0252921100074686 Published online by Cambridge University Press

4. OTHER TOPICS

The previous sections have dealt with fundamentals and deficiencies of aperture synthesis. The main emphasis of this colloquium, however, is to discuss and determine optimum methods of arriving near the true brightness distribution in the presence of imperfections. Only one imperfection is really fundamental to aperture synthesis--the spotty sampling of the visibility function. Others are caused by imperfections in the array and the effect of the troposphere on the radiation.

The exalted, untouchable position of the Fourier solution in equation (2.10) in aperture synthesis has lessened over the last few years for the following reasons: the increased use of high resolution, high frequency observations has prompted a harder look at the general phase problem and its effect on the derived brightness distributions; the sensitive systems now in use enable maps to be made at sensitivity levels for which the desired brightness distribution is comparable or less than the sidelobes of bright sources in the field of view; and *a priori* knowledge about the brightness distribution (positive definiteness, small angular extent, large regions of no emission) are difficult to incorporate into the Fourier method.

4.1. Deconvolution Techniques

The deconvolution technique called clean (Högbom 1974) has been used successfully (with an element of danger) at most radio observatories. The algorithm is able to decompose a radio map into a sum of point-source responses with moderate efficiency, good convergence and now (Schwarz 1978) some understanding of its properties. At the present time, however, very little work has been done on the use of other deconvolution techniques.

The clean algorithm complements the Fourier-inversion solution since the radio map is a true convolution of the apparent brightness distribution and the point-source response. The FFT method and the use of the two-dimensional transform does perturb the exact convolution property; however, these deficiencies can be lessened if necessary.

Clean also works reasonably well for extended sources as long as the loop gain (the percentage of the maximum peak subtracted at each iteration) is set to about 10 percent. Unfortunately, computer time necessary for such cleaning is excessive. Convergence might be faster with more accurate results, if the clean algorithm could deconvolve with a variety of responses of sources with varying angular sizes. First use the broadest beam to clean out the more extended structure and gradually work down to the real point-source response.

https://doi.org/10.1017/S0252921100074686 Published online by Cambridge University Press

4.2. The Phase Problem

The errors of phase of the visibility function have been the main impetus for investigations of non-Fourier reduction methods. By its nature, the standard Fourier inversion treats the amplitude and phase with the same weight and poor maps are generally obtained when non-random phase errors are larger than 20 deg. Other reconstruction methods offer the possibility of using the phase information with less (or no) weight.

The phase problem has long been encountered in VLB observations and optical interferometry. However, until recently these techniques acquired limited sampling of the visibility function and observed preferentially strong sources with relatively small angular diameters, so that the modest technique of model fitting the visibility amplitude was adequate. With increased VLB capability and correlation interferometers now working at optical wavelengths, there is also more pressure to develop efficient algorithms.

Observational and reduction schemes which help reduce the effects of phase errors are

(1) Frequent calibration on an unresolved point source near the radio source
(2) Monitoring the amount of water emission in the direction of the radio source
(3) Use of phase closure to measure accurately the sum of visibility function phase triplets (Fort and Yee 1976)
(4) Use of redundant spacings to separate the effects of the troposphere from the visibility phase
(5) Fourier maps of the square of the visibility amplitude (Baldwin and Warner 1978)
(6) Analytical analysis of the amplitude-spacing data (Bates 1969)
(7) Model fitting the visibility amplitude data
(8) The maximum entropy method
(9) Use of a point source, if any, in the field of view as a phase calibrator.

The major disadvantage to all of the schemes except (1), (2) and (9) is that there must be sufficient signal to noise (> 3 to 1) for *each* visibility function measurement before the techniques work well. Scheme (1) is limited because observations of the calibrator and source are not simultaneous and the calibrator and source are separated in the sky. Scheme (2) does not appear to work very well for a variety of reasons.

Scheme (9), the self calibration of the phase by a point source in the field of view, is promising and has been used in some special cases. The main trick in this method is to isolate the response of the point source by removing the effects of the other emission in the field of view. This might be done in the following way. First, clean the source in the usual way. Next, subtract all of the cleaned components, except the ones associated with the putative calibrator, from the visibility function. With this modified data remake the map which should now include only the point source with sidelobes characteristic of the (u,v) coverage *and* the amplitude and phase fluctuations during the course of the observations. Finally, clean the original map using this new point-source response and iterate.

It is not clear if this particular method would converge to anything useful. This hybrid scheme, I suspect, will be typical of the kind of schemes that will probably be required to make self calibration work.

Closure phase data has been used recently for VLB mapping using a hybrid model-fitting mapping technique (Readhead and Wilkinson 1978). The closure phase probably cannot be used for poor signal-to-noise data because of the 360 degree lobe ambiguities in the sum of the phase of the three visibilities.

4.3. The Three-Dimension Inversion

There is no fundamental problem in aperture synthesis with a non-east/west baseline. The three-dimensional transform of equation (2.9) should be used, if necessary, for skew-baselines. Inadvertent use of the two-dimension inversion would produce 'U'-shaped sources at large angular distances from the phase center. Unlike the bandwidth distortion, however, the convolution property of the map and beam are destroyed in this case.

The w sampling is sparse compared with that in the u and v directions so that the FFT algorithm in all three dimensions would lead to severe aliasing in the n direction. At the VLA the (u,v) inversion will be done using the FFT, the w inversion will be done using the DFT.

4.4. Nomenclature

The nomenclature and symbols used in contemporary aperture synthesis are not particularly uniform, but neither is there much confusion. In this article the term *brightness distribution* has been used for the radio emission I. This is certainly better than the older form of brightness temperature. Perhaps angular power distribution would be more descriptive. There is some confusion in the use of the appropriate sky coordinates. Here and in the general literature (ℓ,m,n) are used for the direction cosines. It is common in radio astronomy to use (x,y,z) except that z usually represents (1-n). For the two-dimensional approximation of the sky, ($\alpha\cos\delta,\delta$) are used. The triad (u,v,w) are commonly used as the ground coordinates associated with aperture synthesis. Here the units of absolute length were used in order to keep the explicit dependence of frequency in the equations. The units of wavelengths are more

https://doi.org/10.1017/S0252921100074686 Published online by Cambridge University Press

commonly used. The trend to use units of delay or time for (u,v,w) is confusing. The direction conversions of (u,v,w) and (ℓ,m,n) given in Section 2.3 are standard with north up and east to the left.

The representation of radio maps is now uniform. Some confusion was caused in the early days when Cambridge displayed maps in which the beams were circularized but the maps were squashed. One bone of contention is the common use of *Jansky per beam area* as an intensity unit. This is the most natural unit from the inversion process, however, some relationship to more externally natural units would be useful. Unfortunately temperature units or Jansky per steradian are not particularly appealing either.

4.5. Map Weighting Terminology

There is confusion among the observatories concerning the meaning of specific weighting (apodizing) schemes in the map making process. *Natural weighting* usually means that the weight of a visibility function sample is proportional to the duration of the sample. The DFT leads to this kind of weighting which gives optimum signal-to-noise for detection of an unresolved source. *Uniform weighting* should mean that the weight of a sample is proportional to the length of the (u,v) track during the period pertaining to the sample. This weighting gives about the highest resolution for the (u,v) coverage. *Smooth weighting* means that the sum of the weights per unit area is reasonably smooth over the entire aperture. This weighting tends to give lowest sidelobe levels. In addition, Gaussian weighting (taper) can be applied.

REFERENCES

Ables, J.G., 1974, Astron. Astrophys. Suppl., 15, 383.
Baldwin, J.E. and Warner, P.J., 1978, Astron. Astrophys. (in press).
Basart, J.P., Miley, G.K. and Clark, B.G., 1970, IEEE Trans., AP-18, 375.
Bates, R.H.T., 1969, Mon. Not. R. Astron. Soc., 142, 413.
Born, M. and Wolf, E., 1964, Principles of Optics, 2nd ed., Oxford: Pergamon.
Brouw, W.N., 1971, Doctoral dissertation, Leiden University.
Brown, R.Hanbury and Twiss, R.Q., 1954, Phil. Mag. (7), 45, 663.
Christiansen, W.N. and Warburton, J.A., 1955, Aust. J. Phys., 8, 474.
Cochran, W.T. et al., Proc. IEEE, 55, 1967.
ERIM, 1977, Study for an Optical Processor System for VLA Radio Telescope System.
Fomalont, E.B., 1973, Proc. IEEE, 61, 1211.
Fomalont, E.B. and Wright, M.C.H., 1974, Galactic and Extragalactic Radio Astronomy, Chap. 10, New York: Springer.
Fort, D.N. and Yee, H.K.C., 1976, Astron. Astrophys., 50, 19.
Glauber, R.J., 1963, Phys. Rev., 130, 2529.
Gordon, R., 1974, IEEE Trans. Nucl. Science, NS-21, 78.

Hamaker, J.P., 1978, Radio Science (in press).
Högbom, J.A., 1974, Astron. Astrophys. Suppl., 15, 417.
Kenderdine, S., 1974, Astron. Astrophys. Suppl., 15, 413.
Lequeux, J., 1962, Ann. Astrophys., 25, 221.
Mandel, L., 1963, Progress in Optics, 2, 181, Amsterdam: North-Holland.
McCready, L.L., Pawsey, J.L. and Payne-Scott, R., 1947, Proc. Roy. Soc., A-190, 357.
Nelder, J.A. and Mead, R., 1965, Computer Journal, 7, 308.
Read, R.B., 1961, IRE Trans. Ant. Prop., AP-9, 31.
Readhead, A.C.S. and Wilkinson, P.N., 1978, Astrophys. J. (in press).
Ryle, M. and Hewish, A., 1960, Mon. Not. R. Astron. Soc., 120, 220.
Schwarz, U.J., 1978, Astron. Astrophys. (in press).
Stanier, H.M., 1950, Nature, 165, 354.
Steel, W.H., 1967, Interferometry, London: Cambridge Univ. Press.

DISCUSSION

Comment L.R.D.'ADDARIO
(1) "Aperture synthesis" is a misnomer, in that the synthesis telescopes in operation today have no equivalent aperture. A better term would be "Fourier synthesis". (Ryle's original use of the term "aperture synthesis" was in a different context.)
(2) The direct Fourier transform should not be abbreviated "DFT", since the latter has long been used in the literature to mean discrete Fourier transform.
(3) I think it is more useful to consider the fundamental equation of synthesis as two dimensional, because (a) the sky is two-dimensional, and (b) measurements in a plane are sufficient. The problem is that we find it impractical to make all our measurements in one plane.

Reply E.B. FOMALONT
I agree with your first two points. I see no way of avoiding an integration in the w-direction for a non-planar sampling of the spatial coherence function in order to obtain an explicit solution for the brightness distribution. For this reason I would consider the 3-D expression more fundamental.

Comment T.W. COLE
The van Cittert-Zernike theorem contains severe restrictions to plane apertures and narrow bandwidths. The fundamental theory can be extended to include the real situation of finite fields of view and finite bandwidths. This is very important now that techniques of point by point Fourier transformation are becoming faster. With such methods compensation is possible for aberrations such as the radial beam smearing. Would you agree that there is still a lot to be done in developing the fundamental theory?

https://doi.org/10.1017/S0252921100074686 Published online by Cambridge University Press

Reply E.B. FOMALONT
I would disagree that there is much to be done in developing the fundamental theory, the van Cittert-Zernike theorem for use in radio astronomy. For example, I believe the radial smearing caused by large bandwidths is unavoidable. I do agree that the smearing is correctable in a variety of ways which I would consider as a deconvolution process.

https://doi.org/10.1017/S0252921100074686 Published online by Cambridge University Press