
Principles of interferometry

Neal Jackson

University of Manchester, Jodrell Bank Observatory
neal.j.jackson@manchester.ac.uk

In this review I outline the principles of interferometry and describe how these are put into practice in a range of modern interferometric telescopes. The basic philosophy of the review is to emphasise a pictorial approach to the subject, rather than delve into detailed mathematical derivations. Much more rigorous treatment of interferometry can be found in Thompson, Moran & Swenson (2001). In addition, the US National Radio Astronomy Observatory (NRAO) hosts a lecture series every three years in interferometry, published in Taylor, Carilli & Perley (1998).

Throughout, I concentrate on applications of interferometry to interferometers at long (wave-regime) wavelengths, with occasional excursions into methods used in the new generation of optical interferometers. The final section is an overview of current interferometers working in the metre-to-millimetre band. The principles behind interferometers at all wavebands, however, are very similar, including all of section 1 and most of section 2. A recent full and complementary review of optical interferometry is given by Monnier (2003), and other reviews in these proceedings cover the VLT interferometer in detail.

1 Basics: Young's slits and Fourier transforms

1.1 Young's slits

Interferometry begins with the Young's slits fringe pattern (Fig. 1). With a single point source emitting coherent radiation, interference fringes are observed, with constructive and destructive interference observed as the relative delay of the two interfering rays changes; the separation of the fringes is λ/d , the wavelength of the light divided by the slit separation.

If the source is made wider (Fig. 1b), we can think of it as a sequence of point sources each of which emit radiation which is uncorrelated with the emission from the others. It follows that the total interference intensity pattern is the sum of the individual patterns. Since an angular displacement in

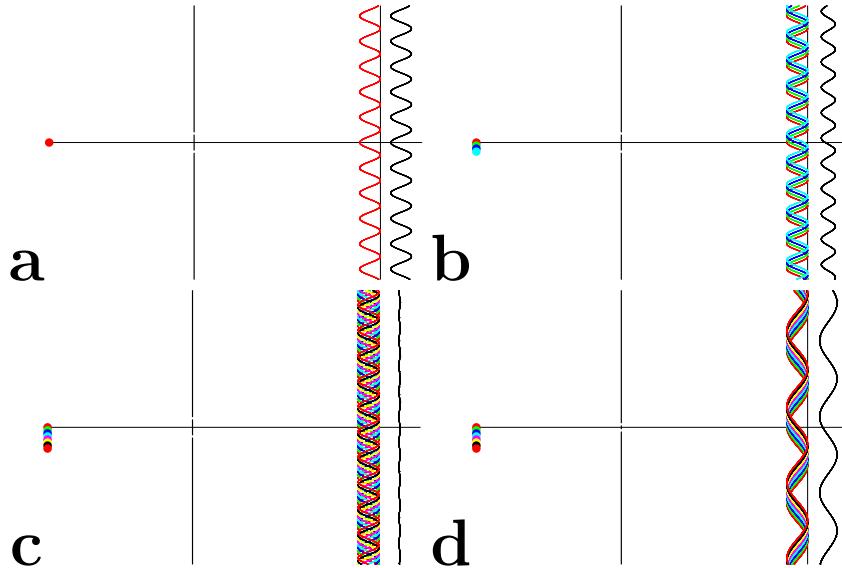


Fig. 1. Young's slits in various situations. In each panel the source is shown on the left, and on the right of the slit are shown the fringe patterns separately for each part of the source and then the added fringe pattern. a: The basic two-slit pattern, showing fringes an angular distance λ/d apart. b: The effect of increasing the source size. An angular shift of the source position by θ shifts the fringe patterns by θ the other way. Since the patterns come from mutually incoherent sources, the intensity patterns add to give a pattern of reduced visibility. c: When the size of the source reaches λ/d , the fringes add to give zero visibility. d: If the slit spacing is then reduced, the fringe spacing increases, and the same size of source is still able to give visible fringes: the source would need to be increased in size to λ/d_{new} in order to wash out the fringes.

the source produces an equal angular displacement in the fringe pattern, as the source size approaches λ/d the fringe patterns will add to give a constant illumination (Fig. 1c). In this case, the fringe visibility (defined as the difference between maximum and minimum intensity, normalized by the sum of maximum and minimum intensity) drops to zero. Conversely, when the angular size of the source is $\ll \lambda/d$, the fringe visibility is 1; this corresponds to a situation in which the source size is smaller than the angular resolution of the interferometer, and only an upper limit of order λ/d can be obtained on it¹.

Now suppose that the slit spacing d is decreased. For the same size of source, this produces less “washing-out” of the fringes, because the same dis-

¹ In practice, the fact that the visibility function begins to decrease as soon as the source extends significantly often allows some information to be derived down to at least $0.5\lambda/d$ and in some cases further.

placement of the source now produces much less displacement of the fringe patterns as a fraction of the fringe separation λ/d (Fig. 1d). The smaller the slit separation, the larger the source size that can be probed using interferometry.

The situation is summarised in Fig. 2. If we plot, for a given source distribution, the way in which visibility varies with slit separation, it can be seen that for small sources the visibility remains high out to large slit separation (in the limit of point sources, to infinite slit separation), whereas large sources produce visibility patterns which fall off quickly as the slit separation increases.

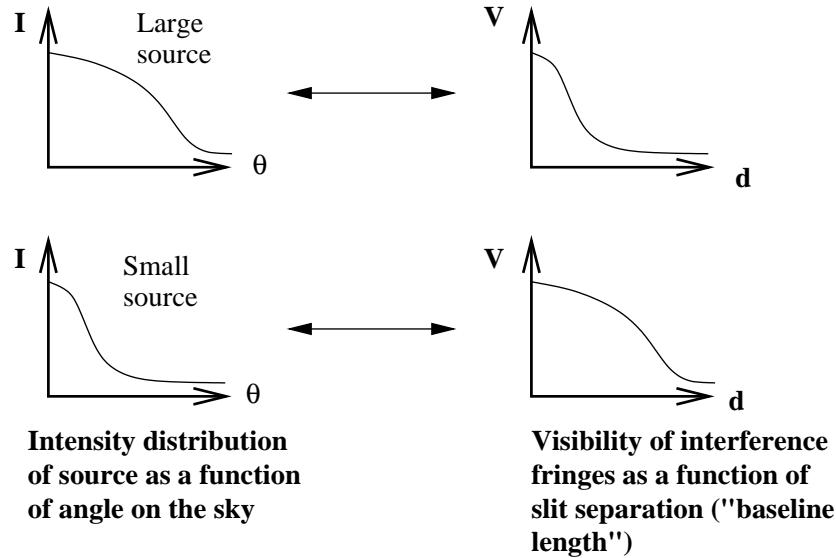


Fig. 2. Relation between source brightness as a function of angular distance and visibility of interference fringes as a function of slit separation (baseline length).

The relation between $I(\theta)$ and $V(d)$ represented here is one which maps a large Gaussian into a small Gaussian, and vice versa, and it is fairly obvious that it is a Fourier transform²; this relationship is the basis of the whole discussion that follows.

This relationship was applied early in the history of radio interferometers to find the sizes of quasars, which were known to be exceedingly small. The method adopted was to use one fixed telescope and one movable telescope, and measure the visibility function using electronic combination of the signals

² This is known as the Van Cittert-Zernicke theorem. Readers requiring a more rigorous derivation are referred to Born & Wolf's (1975) classic textbook on optics, or the introduction to radio astronomy by Burke & Graham-Smith (2002).

over baselines up to ~ 150 km (Adgie et al. 1965). As the telescopes were moved further apart, the visibility finally fell below unity at large separations, allowing the angular size to be calculated as λ/d .

1.2 Application to real interferometers

The Young's slit experiment discussed so far involves sampling two parts of a plane wave generated a large distance away, delaying one wave with respect to the other, and generating the interference pattern as a function of delay. There are many situations in which exactly the same thing is being done. In Fig.3, for example, a plane wave from a source at infinity is sampled by two telescopes separated by a vector \mathbf{B} . The path delay between the two waves is given by $\mathbf{B} \cdot \mathbf{s}$, where \mathbf{s} is the unit vector in the direction of the source, and the phase delay is therefore given by $k\mathbf{B} \cdot \mathbf{s}$, where $k \equiv 2\pi/\lambda$.

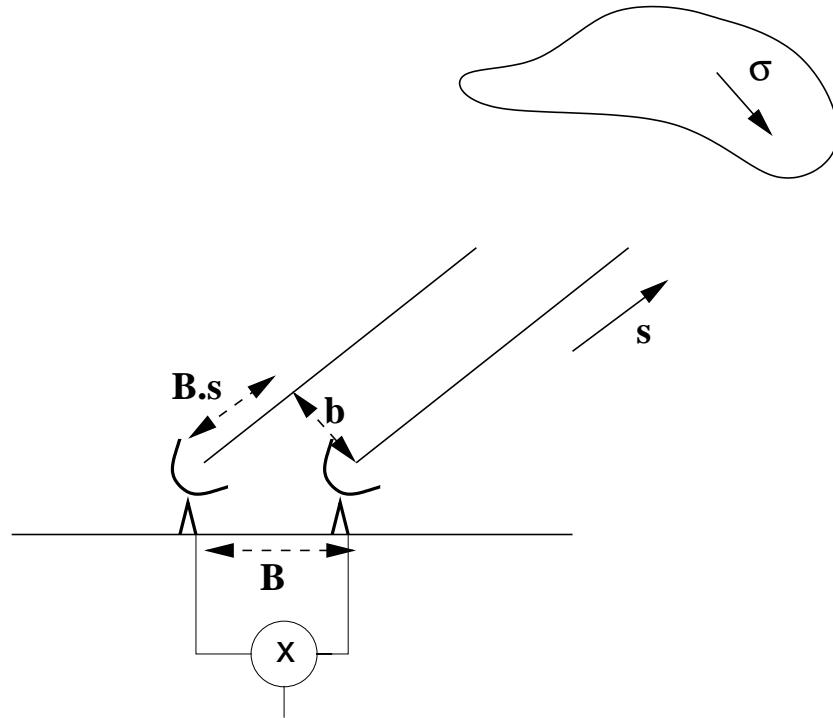


Fig. 3. Basic diagram of an interferometer with baseline vector \mathbf{B} observing a source in a direction with unit vector \mathbf{s} .

Consider a point source. If the electric field received by the first telescope is E , that received by the second is just $Ee^{ik\mathbf{B} \cdot \mathbf{s}}$, because of the phase delay. We can combine these signals, in a way analogous to the screen in Young's

slits, by multiplying them together electronically³ or, in the case of optical systems, by using a Michelson or Fizeau interferometer system to combine the beams. If we then add the fringe patterns over different parts of the source, we obtain the response of the interferometer R as

$$R = \int I(\sigma) e^{ik\mathbf{B} \cdot (\mathbf{s} + \sigma)} d\sigma$$

where $\mathbf{s} + \sigma$ is the vector in the direction of a particular small part of the source with an intensity $I(d\sigma)$. Noting that σ is parallel to the projected baseline vector \mathbf{b} (Fig. 3) and so $\mathbf{B} \cdot \sigma = \mathbf{b} \cdot \sigma$, we then have

$$R = e^{ik\mathbf{B} \cdot \mathbf{s}} \int I(\sigma) e^{ik\mathbf{b} \cdot \sigma} d\sigma,$$

where the $e^{ik\mathbf{B} \cdot \mathbf{s}}$ term is solely dependent on the array geometry and has therefore been removed from the integral.

What we therefore have is a series of fringes, whose amplitude is given by the Fourier transform of the source intensity distribution. In practice, steps are usually taken to get rid of the fringes using a phase rotation whose rate is known (as both \mathbf{B} and \mathbf{s} are known). This is done in optical interferometers by use of accurate delay lines to compensate for the path difference, and in radio interferometers by the insertion of electronic delays. We are then left with the Fourier transform response only, which conveys information about the source. The response is a complex quantity which contains an amplitude and a phase; both are interesting.

Because of the fact that the signal from an interferometer results from the correlation of signals from two telescopes, interferometers have the advantage of much lower sensitivity to interference because most interference does not correlate. Thus the only interference which causes a serious problem is that which saturates or disables the receiver. Such interference can be dealt with by dividing the observing band into spectral channels and removing any channels affected.

1.3 The u,v plane

A further step is to decompose both σ and \mathbf{b} into Cartesian coordinates. The decomposition of σ is easy, as it is just a vector in the sky plane: $\sigma = \sigma_x \mathbf{i} + \sigma_y \mathbf{j}$, where \mathbf{i} and \mathbf{j} are unit vectors in the east-west and north-south directions respectively. This then suggests a decomposition of \mathbf{b} into $u\mathbf{i} + v\mathbf{j}$, so that $\mathbf{b} \cdot \sigma = ux + vy$. The response after fringe stopping then becomes

³ This is not exactly the same as the Young's slits screen, which adds the electric fields and then forms the intensity using $I = (E_1 + E_2)^*(E_1 + E_2)$, but the result is almost the same apart from a constant offset term in the addition case. Not having this term is useful because we do not have to worry about the offset term being constant with time.

$$R(u, v) = \iint I(x, y) e^{2\pi i(ux+vy)} dx dy,$$

a much more explicit 2-D Fourier transform. Note that u and v are defined in units of wavelength, hence the k in the previous expressions has become 2π .

The physical interpretation of the decomposition of \mathbf{b} is fairly straightforward. Imagine sitting on the source (Fig. 4); then the projected baseline vector appears as a line drawn on the earth. This can be decomposed into a component parallel to the equator at its nearest point to the source, and a component parallel to the line between this point and the north pole. These components are u and v , and they change as the earth rotates. Specifically, they trace out an ellipse in u, v space during one earth rotation (Fig. 4).

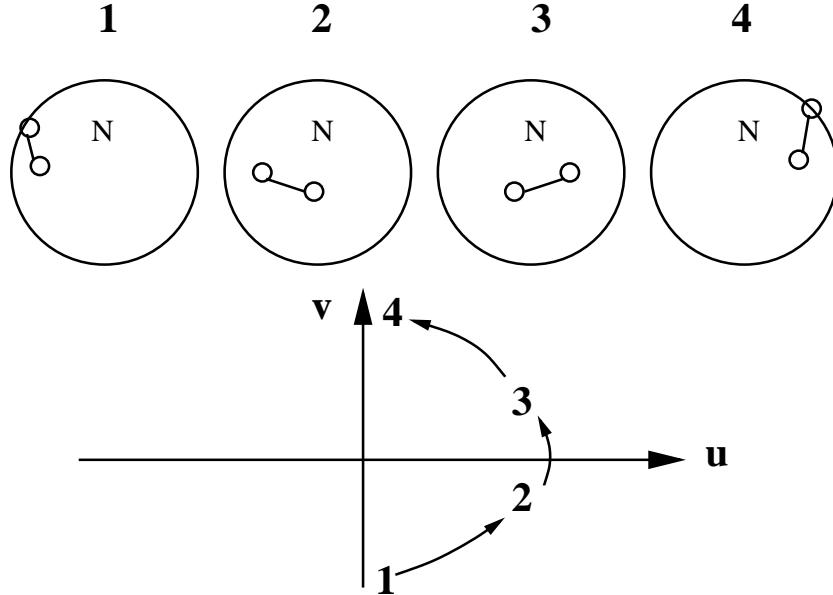


Fig. 4. Schematic diagram showing the baseline between two telescopes as the earth rotates. The E-W and N-S components of this vector give u and v . For an east-west baseline, $v=0$ at source transit.

This change in u and v is useful, as we see if we consider how the response function $R(u, v)$ tells us what is on the sky. Fig. 5 shows the basic Fourier transform relation in a diagram. A double source of separation 1 radian produces stripes in the u, v plane of separation 1 wavelength. Since the Fourier transform gives an inverse relation between distances in the two spaces, a double source of separation a arcseconds gives a series of stripes of separation $206265/a$ wavelengths. Superposed on this is the track of the u, v ellipse, and over a day there are therefore variations in the interferometer response as the interferometer follows the elliptical track over these stripes. Studying

this variation in amplitude (and phase) response over the period, we could work backwards to deduce the separation and orientation of the stripes, and by taking the Fourier transform, recover the source structure.

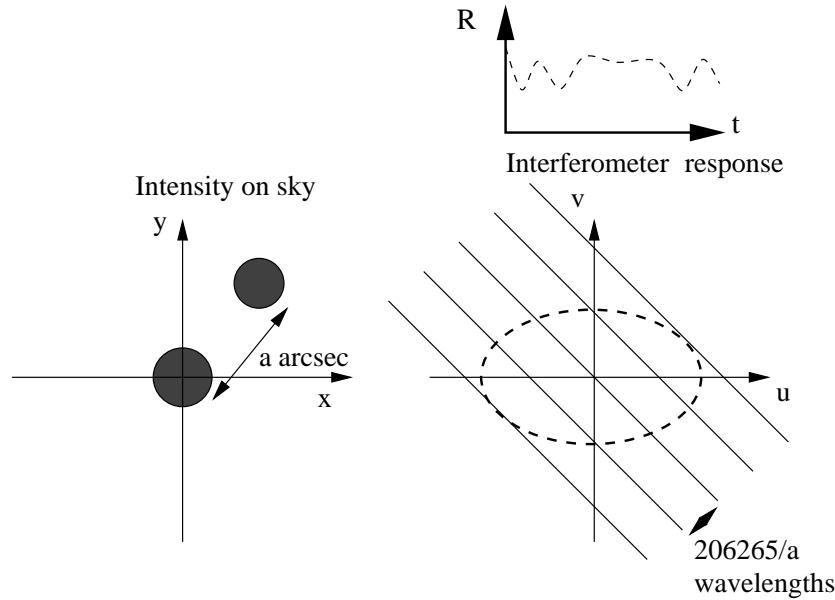


Fig. 5. Diagram showing the interferometer response as a function of u and v for a double source on the sky.

The u, v track has a semimajor axis in the u direction of $\frac{L}{\lambda} \cos \delta$, and a semiminor axis in the v direction of $\frac{L}{\lambda} \cos \delta \sin D$.⁴ In these expressions L is the baseline length, D is the declination of the source, and δ the declination of the baseline; the latter quantity is the declination of the point on the celestial sphere to which the baseline vector points.

The resolution of which a baseline is capable is given by the inverse of the maximum extent of the u, v ellipse, namely λ/L . The point-spread function of an image has dimensions which are the inverse of the spread in the u, v plane of the images being used, which means that for sources close to the equator (where $D \rightarrow 0$ and the ellipses collapse to straight lines) the point-spread function of an interferometer is typically less ideal, although images of the sky can be made with care. Fig. 6 shows the u, v tracks for the MERLIN

⁴ Spherical trigonometry can be used to show that the ellipse is parametrised by the equations: $u = \frac{L}{\lambda} \cos \delta \sin(H - h)$, $v = \frac{L}{\lambda} (\sin \delta \cos D - \cos \delta \sin D \cos(H - h))$ (e.g. Rowson 1963). Here, H is the hour angle of the source and h the hour angle of the point on the sky to which the baseline points. For a general baseline, the centre of the ellipse is offset by $\frac{L}{\lambda} \sin \delta \cos D$ from the origin in the v -direction.

interferometer array, which contains baselines from 6 km to 250 km at a range of orientations. Note the gradual change from circular u, v tracks to nearly linear tracks as the source declination decreases.

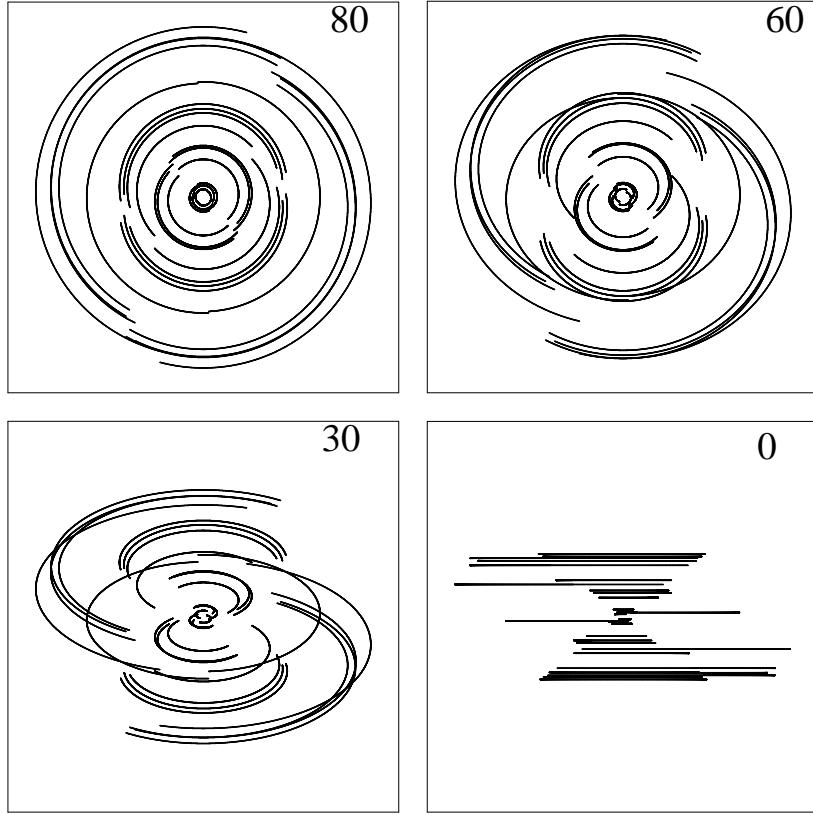


Fig. 6. u, v tracks for the MERLIN interferometer for sources at four different declinations: 80° , 60° , 30° and 0° .

If many baselines are present, many simultaneous measurements can be made in the u, v plane. The more completely the Fourier plane is filled, the easier it is to obtain a faithful reproduction of the sky intensity distribution in an interferometric image.

1.4 A cautionary tale

Interferometric (Fourier) imaging has important differences from direct imaging. The most important difference can be deduced from going back to the Young's slits setup: long baselines record small-scale structure in the source

very well, but are *insensitive* to large-scale structure, because once the source becomes larger than λ/L the fringes wash out and do not return as the source size increases. An example of this is shown in Figs. 7 and 8.

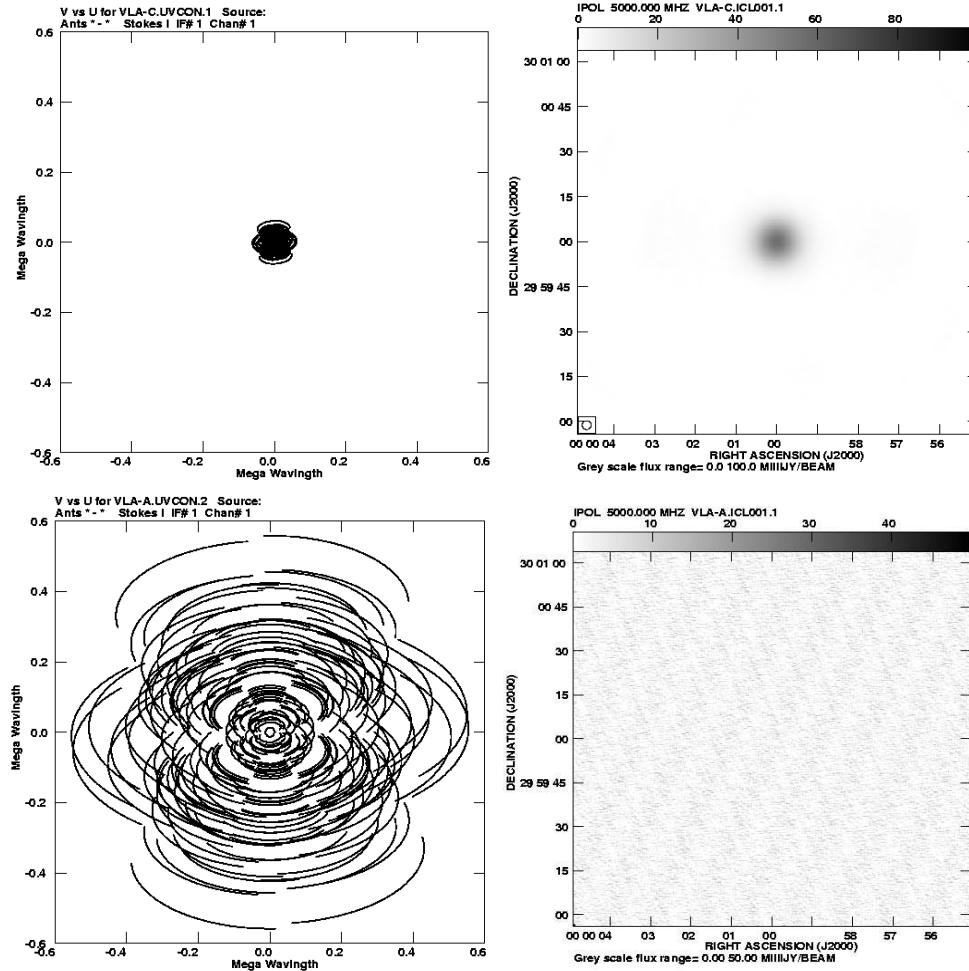


Fig. 7. Simulations of observations of a large Gaussian source with a set of short baselines giving low resolution matched to the size of the source (top left, u, v coverage, top right, resulting image) and a set of long baselines giving higher resolution (bottom left, u, v coverage, bottom right, resulting image).

Fig. 7 (top panels) shows a simulated source of $12''$ extent, mapped using an array whose u, v coverage gives a maximum baseline length corresponding to a resolution of $3''$. The source structure is recovered reasonably well, as

the range of baselines present cover spatial scales from the resolution up to approximately 10 times larger scales. Now suppose that we get greedy, and decide that we would like to observe at ten times higher resolution. This is no problem – we move the telescopes to spacings ten times greater and repeat the observation. This indeed gives a map with 300-mas resolution, on which there is no sign of the source (Fig. 7, bottom panels).

Let us then try to smooth the image and recover the structure. If we try this, what we actually recover is shown in Fig. 8. The awful realisation dawns at this point that by using a long set of baselines, we have not recorded the structure on large spatial scales at all, and have lost it irretrievably. The moral is that interferometer arrays should be chosen carefully to match resolution to the spatial scales required by any particular astrophysical problem⁵.

1.5 Field of view of interferometric images

Primary beam

Once again, we can go back to Young’s double slits to deduce the another fundamental limitation of the interferometric image. If the slits are widened, the aperture distribution no longer consists of two delta-functions, but of two delta functions convolved with a single wide slit. It follows from the convolution theorem that the interference pattern, being the Fourier transform of the aperture distribution, consists of the original two-slit fringe pattern multiplied by the Fourier transform of a wide slit, namely a sinc function. The sinc function has a width inversely proportional to the width (w) of the slits, and the fringes disappear at delays greater than the width of the sinc function, λ/w .

Now in an interferometer, going further away from the centre of the field of view just corresponds to a different delay from that which obtains at the centre. The width of the slits translates directly to the size of each interferometer element, and the field of view in radians is the wavelength of the light being studied divided by the diameter of the elements.

Wavelength ranges

Again going back to Young’s slits, it is easy to see that other effects may intervene before the primary beam limit is reached. The most serious of these

⁵ In fact, using resolution higher than required often causes even worse problems. This is because for any given array, higher resolution demands observing at higher frequencies, which in turn imposes penalties in source brightness for typical steep-spectrum radio sources and in generally worse system performance at high frequencies. Some interferometers, in particular the VLA, allow different resolution images at the same frequency by regularly moving the telescopes between different configurations, from compact to more extended.

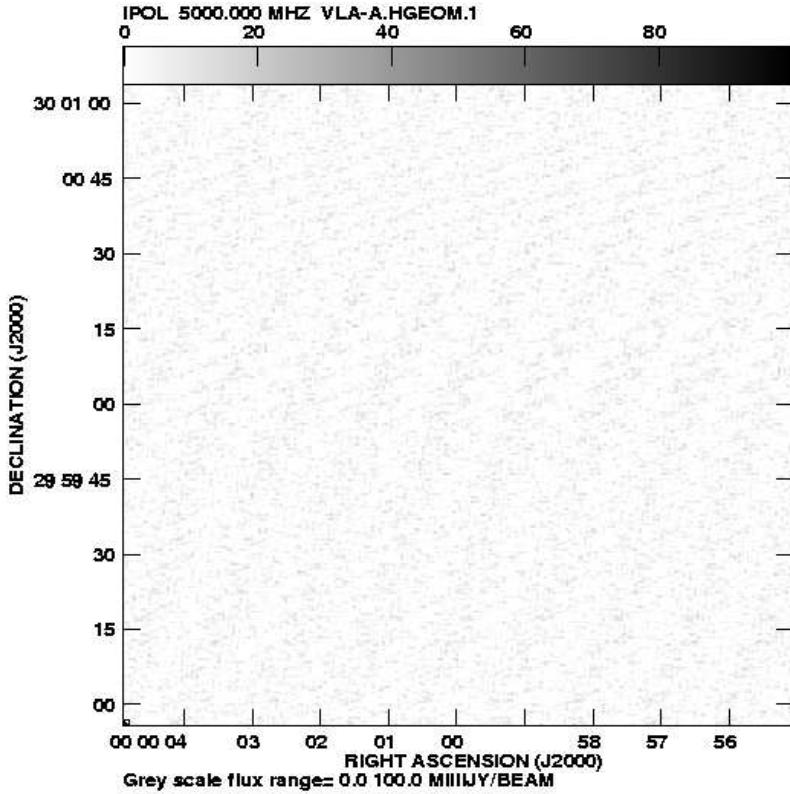


Fig. 8. Smoothed image of a long-baseline observation of a diffuse source. Nothing is visible.

is that the radiation is not monochromatic. We can consider each single frequency separately, and add the resulting fringe patterns which have different separation (λ/d) between maxima. The result is that in the centre of the fringe pattern, full-visibility fringes are seen, since here the delay is zero. At larger values of delay, further up the screen, the interference fringes from different colours add in such a way as to reduce the visibility to zero even for a point source. Once again, the effects at large delay translate directly to the interferometer, and a range of wavelengths in the interferometer causes loss of response at the edge of a field of view where the delay between the interfering waves is different⁶. If the bandwidth is $\Delta\lambda$, then the field of view is given by

⁶ For a large Gaussian-shaped bandwidth, the fringe response as a function of angular distance from the centre is a small Gaussian, and for a small Gaussian-shaped bandwidth, the fringe response varies as a broad Gaussian with angular distance. This completes another “proof” of a Fourier transform relation between these two quantities.

$$\Delta\theta = (\lambda/\Delta\lambda)(\lambda/L),$$

or the beam-size divided by the fractional bandwidth. In order to achieve reasonable signal-to-noise, many interferometers use large fractional bandwidth, implying a very restricted field of view. The solution is simply to divide the signal into many frequency channels and correlate each separately. The cost is greater complexity of the correlator and larger datasets, but in most modern interferometer systems computing and hardware are advanced enough that it is usually possible to image the full primary beam.

Other effects

Two other effects should be mentioned briefly; see Taylor et al. (1998) for further details. The first is that a limitation on the field of view is imposed by the integration time per data point, because the values of u and v change during a finite integration time. This gives a roughly tangential smearing in the u, v plane which becomes worse further out, which Fourier transforms into a tangential smearing in the sky plane which becomes worse further out. The result is that amplitude is lost at the edge of the field. The second effect is that the sky is not flat, and that instead of using a 2-D Fourier transform we should have used a 3-D transform, with an extra phase term of the form $\sqrt{1 - x^2 - y^2}$ added to the transform. Unlike the bandwidth and integration time effects, the effect of non-flatness is curable after the event by additional processing.

2 Producing the image

2.1 Deconvolution

So far we have circumvented the major problem, which is that the interferometer response function has not been measured over the whole u, v plane. To do this at a single frequency would require enough telescopes to provide baselines at all possible separations and orientations, an expensive operation with substantial planning implications. Lack of this information means that the number of different images consistent with the data is infinite, since we could in principle fill in the unmeasured parts of the u, v plane in an infinite number of ways.

The basic problem is that we want the image $I(x, y)$ resulting from the full u, v response function $I(u, v)$,

$$I(x, y) = \iint I(u, v) e^{2\pi i(ux+vy)} dudv$$

but instead have the “dirty image”

$$I_D(x, y) = \iint I(u, v)S(u, v)e^{2\pi i(ux+vy)}dudv$$

which results from the intervention of the sampling function $S(u, v)$ which is 1 in parts of the u, v plane where we have sampled and zero where we haven't.

We recognise the right-hand side of the last equation as a Fourier transform, where the argument is the product of two functions I and S . We can therefore use the convolution theorem to write

$$I_D(x, y) = I(x, y) * B(x, y)$$

where

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

the “dirty beam”, is the Fourier transform of the sampling function. Since we know where the telescopes are and can do spherical trigonometry and Fourier transforms, the sampling function and hence dirty beam are accurately known. Recovering the image $I(x, y)$ is therefore a classical deconvolution problem, in which we need to supply additional information in order to do the deconvolution.

CLEAN

The first way to do this is the algorithm known as CLEAN (Hogbom 1974) which amounts to a brute force deconvolution. The basic algorithm begins by detecting the brightest point in the dirty map, shifting the dirty beam to this point, and scaling and subtracting off the dirty beam⁷. At each subtraction, the flux and position subtracted are noted, until the map from which the dirty beams have been subtracted (known as the residual map) consists only of noise. At this point, the subtracted fluxes are convolved with a restoring beam selected by the user and added back into the field of noise to give a final “CLEAN map” from which the sidelobes of the dirty beam have been removed. The usual procedure is to make the CLEAN beam of the same dimensions as the central spike in the dirty beam. This is a logical procedure, since the dirty beam is the Fourier transform of the sampling function, and the further out in the u, v plane the sampling function has non-zero values, the higher the resolution we are justified in using. It is possible to use a CLEAN beam smaller than the formal resolution (a process known as super-resolution) at increasing risk of introducing incorrect structure in the map. Fig. 9 shows an example of the CLEAN procedure in action.

An important decision in the CLEANing process is the weighting to be applied to the data, as the recorded data is not uniformly distributed across

⁷ In practice, a fraction – typically 5–10% – of the dirty beam is subtracted to improve stability. This fraction is known as the ‘loop gain’.

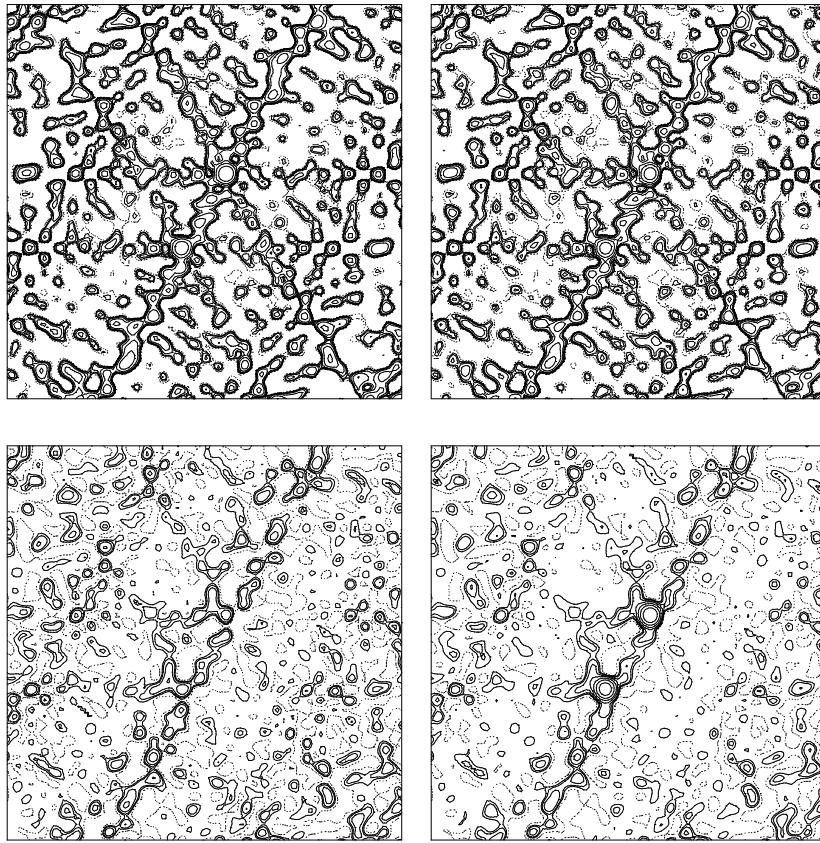


Fig. 9. CLEANing procedure applied to a radio source consisting of two point source components observed with the VLA. All maps are contoured at the same level. Top left: the dirty map. Top right: the residual map after 10 iterations of CLEAN, in which a small amount of flux has been removed at each iteration. Bottom left: the residual map after 100 iterations of CLEAN: note the removal of most of the dirty beam structure. Bottom right: the CLEAN map after some further CLEANing, formed by the addition of the point source components back into the final residual map. Further CLEANing does not give significant improvement; although the basic source structure is visible, there are some clear artefacts remaining. Their causes and cure are addressed in section 3.

the u, v plane and in practice is usually concentrated towards the centre. One option is “natural weighting”, in which all data points are treated equally. Statistically, this provides the best signal-to-noise in the final image, but because of the central concentration of the u, v data the sampling function is more centrally concentrated, and its Fourier transform, the dirty beam in the sky plane, is therefore more extended. The result is worse resolution in the final map. An alternative option is “uniform weighting” in which equal weights are applied to each u, v grid, giving increased resolution at the expense of weighting down data at small u and v and degrading the signal-to-noise.

The additional information that has been supplied to the deconvolution problem by CLEAN is the assumption that the sky consists of a finite number of point sources, or alternatively that most of the sky is empty. Not surprisingly, therefore, CLEAN works very well for simple sources, but can occasionally fail on very large amorphous sources of low surface brightness.

Maximum Entropy

A second deconvolution method is altogether different in philosophy, and is known as the Maximum Entropy Method, or MEM (e.g. Bryan & Skilling 1980). The starting point is to consider possible images of the sky, and prefer those which are more likely. The most preferred image is a completely uniform distribution, which gives the maximum entropy (or minimum information); however, such an image is normally inconsistent with the data. If we consider images which are progressively less likely to be produced by chance, sooner or later we encounter an image which still occurs relatively often, but which is nevertheless consistent with the data. The process therefore corresponds to a joint minimization including the goodness of fit to the data and the maximum effective smoothness (usually parametrised in forms such as $\sum p_i \ln p_i$, where the p_i 's are the individual pixel values).

2.2 Sensitivity

The sensitivity (r.m.s. noise) of a wave-regime (metre-centimetre) interferometer is given by

$$S = \frac{\sqrt{2}k_B T_{\text{sys}}}{A\eta\sqrt{n_b}\Delta\nu t_{\text{int}}}$$

where T_{sys} is the system temperature, A is the area of each antenna, η is the aperture efficiency, n_b is the number of baselines, $\Delta\nu$ is the observing bandwidth and t_{int} is the integration time. The units are $\text{Wm}^{-2}\text{Hz}^{-1}$, but because the Boltzmann constant k_B is uncomfortably small the usual unit is the Jansky, where $1 \text{ Jy} \equiv 10^{-26} \text{ Wm}^{-2}\text{Hz}^{-2}$. For extended sources the sensitivity is Janskys per beam area. The sensitivity of many modern interferometers after a few minutes of integration is around $100\mu\text{Jy}/\text{beam}$.

A couple of terms in the equation deserve comment. We can define “temperatures” in this context in terms of the temperature of a black body which would provide the equivalent received power of radiation at the observing frequency⁸. The noise contribution to a radio interferometer is provided mainly by the receivers (contributing typically 30-50K), spillover from thermal emission from the ground, and ultimately by the 3K contribution from the cosmic microwave background. The aperture efficiency η can be varied according to how the aperture is illuminated (i.e. the relative weight given to radiation reaching the feed from different parts of the antenna).

In an optical interferometer the formulas are somewhat different, because we are collecting photons. There are a number of detailed differences. The first is that it is impossible to clone photons in the same way that electrical signals can be reproduced indefinitely, so every time a beam is split, signal-to-noise is lost. For example, if the array consists of eleven elements, the beam from each element must be split ten times, losing signal-to-noise, if we wish to interfere all of the beams to produce fringes on all baselines. Secondly, the practical limit is always imposed by the fact that we need a reasonable number of photons in one isoplanactic patch (the area over which the atmospheric corruption is approximately the same) in one atmospheric coherence time (the timescale of variation of atmospheric corruption). Third, optical interferometers typically contain a large number of reflecting elements, with some light loss at each reflection.

Although it is probably fair to say that the problems have been more difficult than anticipated, major progress is now being made. The use of adaptive optics on individual telescopes means that the wavefront can be corrected over the whole aperture, increasing the coherence patch to the area of the telescope diameter and hence increasing the potential signal. In an optimum site such as that of the Very Large Telescope Interferometer (VLTI) on Paranal mountain in Chile, images of 14th magnitude objects can be made. A list of current optical interferometer systems is given by Monnier (2003).

3 Dealing with the atmosphere

Electromagnetic radiation travels to us for billions of years through a nearly perfect vacuum as a nearly perfect plane wave. Unfortunately, the Earth’s atmosphere intervenes in the last microsecond to convert a smooth wavefront into a wavefront with phase corrugations which vary over small spatial scales and on potentially small timescales. At some wavelengths, the amplitude of the wavefront is affected as well as the phase.

There are a number of features in the earth’s atmosphere which corrupt the wavefronts. At low radio frequencies, the problem is the ionosphere, con-

⁸ In the wave regime, we are in the Rayleigh-Jeans part of the Planck spectrum, and the specific intensity at a given frequency (in units of $\text{Wm}^{-2}\text{Hz}^{-1}\text{sr}^{-1}$) can be written as $2k_B T/\lambda^2$

sisting of a collection of charged particles capable of shifting phases below the plasma frequency (typically a few hundred MHz) and which responds to solar activity, producing most disturbance at times of solar maximum. At higher radio frequencies, the problems are mainly due to water vapour, which produces phase rotations on “coherence timescales” of minutes above 10 GHz. Successively shorter coherence times are seen as the frequency increases, until not only the phase but also the amplitude is affected. Observations at higher radio frequencies, such as the ≥ 30 GHz observations typically used to observe the peak of the CMB radiation, are usually done from high mountains above most of the water vapour; some experiments in this region are planned at the South Pole where the water vapour is frozen out.

In the infrared and optical region of the electromagnetic spectrum, the coherence times are typically much shorter. The transverse length scale of phase fluctuations is given by the Fried parameter r_0 , and these atmospheric fluctuations are blown across any given line of sight by tropospheric winds. The resulting phase and amplitude fluctuations have characteristic timescales of tens of milliseconds, requiring corrections to be applied on short timescales which are now being achieved.

3.1 Closure quantities

Ignoring atmospheric phase fluctuations is not an option, as the response function of the interferometer is directly affected by them and their effect is to wipe out the fringes. The phase on any individual baseline is not a good observable, since atmospheric errors e_1 and e_2 on two telescopes produce a resultant of the form $e_1 - e_2$ in the response function when the signals are correlated⁹.

We can observe, however, that if we have a triangle of three telescopes, and if we measure the interferometer phase response on each baseline, we obtain three phases responses containing the phase error terms $e_1 - e_2$, $e_2 - e_3$ and $e_3 - e_1$. These add to zero, leaving only information on the astronomical structure. Although we have slightly fewer constraints, by modelling these “closure phases” we can obtain constraints on the phases of the response function in the u, v plane and hence deduce the source structure.

A similar quantity can be derived for amplitudes. Since amplitudes are multiplicative, four telescopes are needed in order to use the baseline amplitudes to form the quantity $A_{12}A_{34}/A_{13}A_{24}$. The error on A_{12} is the product of the amplitude error terms, a_1a_2 , and once again the errors cancel out.

Closure mapping has been used for many years. It has successfully been used in optical interferometry, even from sites with considerable phase fluctuations such as the situation of the COAST optical interferometer. Despite the less than ideal atmospheric conditions, maps have been produced of bright

⁹ Correlation involves an operation of the form $\langle E_1 E_2^* \rangle = A_1 e^{i\phi_1} A_2 e^{-i\phi_2}$, so amplitudes multiply and phases add or subtract.

stars such as Capella and Betelgeuse (Young et al. 2004). This gives a clue to the main limitation, however; use of this method requires that the sources be bright. This is because what is required is sufficient signal on the source in an atmospheric coherence time to separate atmospheric and source phases.

3.2 Self-calibration

The assumption of closure mapping is that errors in amplitude and phase are separable by telescope, and that no additional errors are introduced into the response function of each baseline separately. This is not precisely true, and in practice the biggest problem is usually mismatched bandpasses in the correlator which gives baseline-dependent errors. Great care is usually taken to minimise these, resulting in such errors being a few tenths of a percent.

The assumption of only telescope-based error is used in the procedure known as self-calibration (Cornwell & Wilkinson 1981) which uses the data, together with a guessed model, to determine the phase and amplitude corrections on each telescope. Suppose we have visibilities V_{ij} on baselines between telescopes i and j , and we call the telescope complex gains g_i and g_j . Suppose also that we have a model whose Fourier transform predicts visibilities V_{ij}^M . Then we write the equation

$$V_{ij} = g_i g_j V_{ij}^M$$

for all i, j , and use a least-squares solution to determine the g_i s. The process is repeated by replacing the original data with $V_{ij}/g_i g_j$, mapping and CLEANing the new data to produce a new model, Fourier transforming to give a new set of model visibilities $V_{ij}^{M'}$, and repeating the process until it converges and the g_i corrections are close to 1.

At first sight this looks an uncomfortably incestuous procedure. A model which may or may not look like the sky has been used to correct the visibility data, and we have then used a model derived from the corrected data to determine further corrections. One answer is that the procedure works. Simulated data can be created, phase and amplitude errors added and the original sky map is recovered by self-calibration. The underlying reason for its success is that the problem is overconstrained, because in any one integration the number of unknowns is proportional to the number of telescopes, n , and the number of constraints is proportional to the larger number of baselines, $n(n - 1)/2$. Fig. 10 shows an example of self-calibration in action.

There are a number of caveats in practice. The most important is that the source needs to contain a point source bright enough to be visible on all baselines at $> 3\sigma$ in one coherence time – for many interferometer arrays, this means of the order of 10–20 mJy. The reason is that the least-squares fit in which the g_i s are computed degenerates into a noisy mess if the V_{ij} s are noisy. Because the corrections change over a coherence time, it is not possible to integrate for a long time in order to build up signal-to-noise to do the correction.

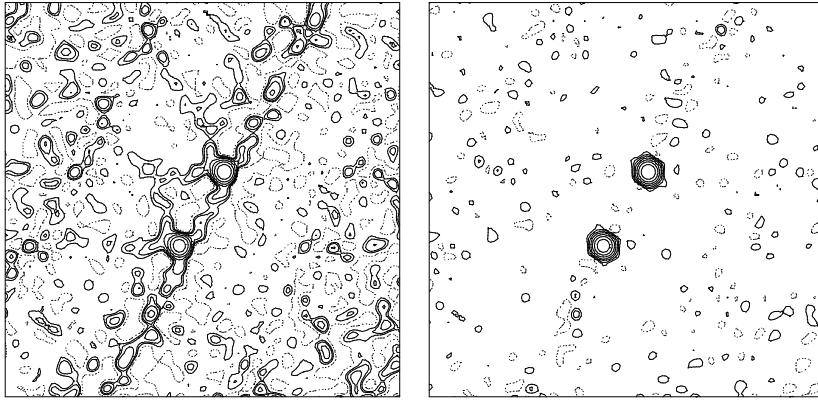


Fig. 10. The same radio source as in Figure 9. The left panel shows the deconvolved map after CLEAN only. On the right is the same map after one iteration of phase selfcalibration and further CLEANing. The maps are contoured at the same level, but the impact of selfcalibration in removing artefacts due to phase corruption is obvious.

A second caution concerns the order in which the phase and amplitude corrections are built up. Since atmospheric effects on phase are nearly always more major than amplitude effects, it is usually better to begin by computing only the phase part of the complex gains, using a time interval shorter than the phase coherence time, and only then to correct the amplitude part. This halves the number of free parameters in the early part of the process and thereby makes it much more stable. Once the phases are determined, the amplitudes can be corrected over a longer timescale as (at least at a few GHz) the amplitude corruption does not change as quickly. Indeed it is often a good idea, particularly in sparsely filled arrays or for relatively weak sources, not to use too short a correction time for the amplitudes. For most VLBI experiments 15-30 minutes is probably safe.

3.3 Phase calibration

A more direct way to correct the atmospheric phase corruption is to calibrate it directly, rather than sort it out after the event. In this approach, a calibrator source of known structure is observed periodically. Because the source structure is known, it can be Fourier transformed and removed from the interferometer response function. Any residual phase structure must be atmospheric and can be interpolated and removed from observations of the target.

This approach is attractive because, provided the calibrator source is strong enough to allow good signal-to-noise per baseline per atmospheric coherence time, there is in principle no limit on the brightness of the target

source. Phase calibration, otherwise known as “phase referencing”, is therefore widely used. There are two caveats: first, the target source must be in the same isoplanactic patch (that is, the target and phase calibrator must be close enough that the atmospheric phase corruption is similar for both), and the switching must be done with a period not greater than the atmospheric coherence time.

4 Interferometers in practice

4.1 Very brief history

Much of the development of connected-element interferometry was done by groups in the UK, USA, the Netherlands and Australia, and a Nobel Prize was awarded to Martin Ryle in Cambridge for development of the technique. This group used arrays of dipoles, and later connected dish antennas together forming the One-Mile Telescope, to discover and investigate radio sources (Ryle 1962). The earliest large radio source catalogue, 3C, contained the first two known quasars, 3C48 and 3C273. The successor to the One-Mile Telescope, the 5-km Telescope (now known as the Ryle Telescope, Ryle 1972) was built soon afterwards.

At the same time, other interferometer arrays were being built, including the Westerbork Synthesis Radio Telescope (WSRT) in the Netherlands (Allen, Hamaker & Wellington 1974), the Bologna Cross, arrays at Molonglo in Australia and Ooty in India. The MERLIN six-telescope interferometer array, based at Jodrell Bank, pioneered the extension of connected-element interferometers to longer baselines of around 200 km giving higher resolution (Davies et al. 1980). Also in the late 1970s, the Very Large Array (VLA, Thompson et al. 1980) was built. This instrument has 27 25-m diameter telescopes, giving high sensitivity, and has a maximum baseline of 36 km. It is arranged in a Y-shape, and the antennas can be moved to four different configurations with baselines shorter by successive factors of 3.

4.2 More current and future interferometer systems

A brief overview of some long-wavelength interferometer systems follows (see also Fig. 11). Again, the new optical systems, such as the VLTI, COAST and Keck interferometers are covered by the article by Monnier (2003).

VLBI

An important subset of interferometer arrays are those operating on very long baselines, known as VLBI (very long baseline interferometer) arrays. The original such collaboration was the European VLBI Network (EVN) whose founding member telescopes included the 76-m Jodrell Bank telescope in the UK,

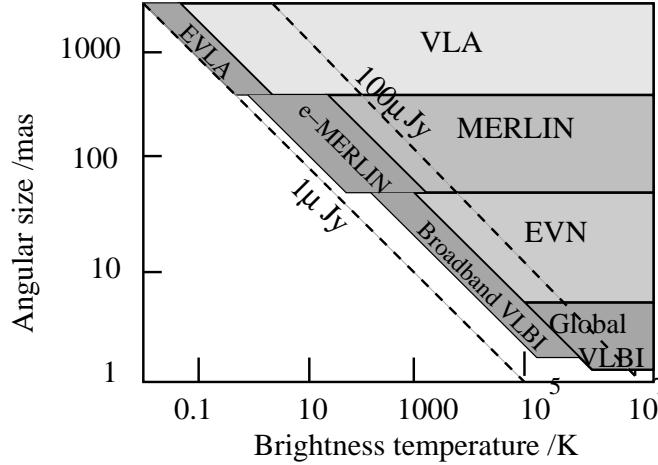


Fig. 11. Range of angular source size (set by the resolution) and brightness temperature of sources, with the northern interferometer array suited to the observation at 5 GHz. The lower limit to each range is set by resolution, and the upper limit by the insensitivity of interferometers to sources larger than that visible to their shortest baselines. The upgrades described in the text are included in the figure. The strength of the sources is plotted as “brightness temperature” which is related to flux density S in Jy by the equation $S = 2k_B T \Omega / \lambda^2$, where Ω is the beam solid angle. Reproduced from the MERLIN website www.jb.man.ac.uk/~merlin.

the WSRT in the Netherlands¹⁰, the 100-m Effelsberg antenna near Bonn, Germany, the Onsala 25-m telescope in Sweden and the 32-m Bologna telescope in Italy. Further telescopes now part of the EVN include the Torun telescope in Poland, and further antennas in Italy, Spain and China. A 10-telescope array, the Very Long Baseline Array (VLBA) was subsequently built in the USA with somewhat higher resolution but lower sensitivity. It is possible to combine the EVN and VLBA into a global VLBI array with high sensitivity (about $10\mu\text{Jy}/\text{beam}$ after 12 hours) and a resolution of 1 mas, about 50 times higher than that of the Hubble Space Telescope. It is also possible to increase the baseline still further by launching a radio antenna into space. This was done experimentally by use of the TDRSS satellite (Linfield et al. 1989) and later by the dedicated VLBI satellite VSOP/Halca, launched by the Japanese space agency in 1997 and which observed until 2003. A larger mission (VSOP-2) is currently funded, which will feature a larger telescope to be launched in 2012.

¹⁰ The WSRT is itself a connected-element interferometer, but it is possible to insert phase delays into the arm of each element in such a way as to use the interferometer as a single telescope with effective area of the sum of the individual telescope areas (a process known as “phasing up”).

VLBI observations are usually not combined into visibilities at the time of observation. Instead, they are often recorded on tape, together with accurate time stamps from a maser clock, and shipped to a central processor for correlating, usually the JIVE facility in the Netherlands or the NRAO correlator in Socorro. Recently, with the availability of increased Internet bandwidths, it has been possible to send the signals to the central correlator online, a process known as e-VLBI. In principle, the only limit on this technique is the speed of the correlator and the availability of the huge Internet bandwidths required.

GMRT

The Giant Metre-wave Radio Telescope (Swarup et al. 1991) is located in India, near the city of Pune. It consists of 30 45-m antennas, and the resulting large collecting area gives very high sensitivity between 50MHz and 1420MHz; the frequency range is limited at the high end by the fact that the antennas are of mesh rather than solid metal. It is particularly suitable for relatively high-resolution, sensitive imaging at low frequencies including the redshifted neutral hydrogen line.

ATCA

The Australia Telescope Compact Array is the most significant long-wavelength interferometer system in the southern hemisphere. It consists of six 22-m antennas over baselines of up to 6 km, with good high-frequency performance.

Fibres and sensitivity: EVLA and e-MERLIN

Important upgrades to Earth-based interferometers are currently under way. The major programmes involve an increase in sensitivity by using higher bandwidth. Currently signals are transmitted using transmission lines or microwave links with a limited bandwidth, typically a few tens of MHz. Both MERLIN and the VLA are being upgraded within the next few years (becoming, respectively, e-MERLIN and the EVLA) by the addition of optical fibre links between telescopes. These links can carry signals of \sim 2GHz of bandwidth, resulting in a factor of 5–10 in improvement in signal-to-noise, and in both cases for a small fraction of the cost (in current dollars) of the original arrays. The increased bandwidth has another benefit when observing a broadband source, namely an improvement in u, v plane coverage. Since the u, v plane is measured in wavelengths, a wide band means that for a given baseline, a range of positions in the u, v plane are measured simultaneously. This means that even an array with a small number of elements, such as MERLIN with six telescopes, can cover essentially all of the u, v plane and thus deliver very high image fidelity (Fig. 12). With complete aperture coverage, radio interferometers will be able to produce the same level of detail as in current direct optical images.

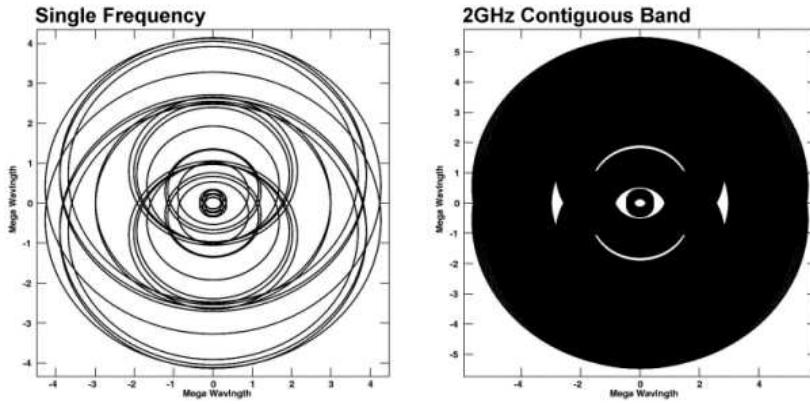


Fig. 12. The difference between u, v coverage of MERLIN, a six-telescope interferometer array, using a single frequency (left) and a 2-GHz bandwidth at 5 GHz (right). The essentially complete coverage of the u, v plane allows extremely high fidelity imaging of much more complex structure than hitherto possible. The processing power needed is considerably greater, because of possible spectral index gradients across the source, but this too is essentially a solved problem. The image is reproduced from the e-MERLIN science case.

LOFAR and the MWA

At the low-frequency end of the radio spectrum, it is possible to increase sensitivity using the fact that huge collecting areas can be achieved relatively cheaply. This is being exploited by the Low Frequency Array (LOFAR, Röttgering 2005) being built in the Netherlands, and by the Mileura Widefield Array (MWA) currently undergoing demonstrator tests in Western Australia. LOFAR will consist initially of about 50 elements, each consisting of a field of 50-m diameter filled with dipole antennas. This will give high sensitivity up to 240MHz and a resolution of a few arcseconds, with a possibility of higher resolution if long baselines are added. At such low frequencies, the field of view is very wide and survey speeds are correspondingly high. It is also possible to manipulate the phases applied to the antennas to form multiple beams on the sky, effectively allowing the telescope to look in a number of directions at once. The major science goals include the detection of the epoch of reionization in redshifted neutral hydrogen, the production of sensitive wide-field surveys and the monitoring of transient sources which becomes possible with rapid sky coverage¹¹. The difficulties include the large amount of process-

¹¹ Another widefield telescope at somewhat higher frequencies is the Allen Telescope Array, currently being built at a site in New Mexico, USA. When complete it will consist of 300 6-m dishes and have a large survey speed by virtue of the sensitivity from the large number of elements combined with the large primary beam of the small individual elements.

ing power required, wide-field problems with removing sidelobes from bright sources at large angles, and more seriously the calibration problems associated with dealing with rapidly varying phase corruption from the ionosphere and (particularly in the LOFAR case) the necessity for very efficient excision of radio-frequency interference.

ALMA

At the other end of the frequency range, the Atacama Large Millimetre Array (ALMA) is a new interferometer being built on the Chajnantor plateau in Chile, a dry and high (5000-m) site close to the Bolivian border. The choice of site is due to the effects previously mentioned of the water vapour in the lower atmosphere on the amplitude of radio signals. ALMA will operate between 30 and 950GHz, in the windows permitted by the small quantity of atmospheric water vapour which remains above it. Its strength lies in the wide variety of molecular astrophysics and chemistry which can be probed at these frequencies, due to the huge number of molecular lines in the millimetre and submillimetre bands; it will be able to probe areas of star formation very sensitively, as well as detecting molecular gas on cosmological scales from distant galaxies.

CMB observations

A specialised niche in interferometry is occupied by experiments which are detecting structure in the Cosmic Microwave Background. These experiments tend to use short baselines, because the power in the CMB fluctuations is observed at significant strength on scales from arcminutes up to degrees, and frequencies from a few tens of GHz upwards due to the fact that the CMB radiation has a thermal spectrum with a temperature of 3K and a consequent spectral peak at 300 GHz. Most, like ALMA, operate from high, dry sites such as the Chajnantor plateau itself, the island of Tenerife or the South Pole.

Square Kilometre Array

After the bandwidth of interferometers has been increased to a maximum ($\Delta\nu/\nu \sim 1$) there is only one option for increasing sensitivity, namely increasing the collecting area. In the wave regime, the noise level becomes better as A^{-1} rather than the $A^{-1/2}$ obtained by increasing the collecting area of a photon collector, so increased telescope acreage is rewarded particularly spectacularly in the low-frequency part of the electromagnetic spectrum.

The idea of a very large interferometric telescope arose fifteen years ago (e.g. Wilkinson 1991) and is now under detailed design study with a view to full-scale construction in either western Australia or South Africa in the middle part of the next decade. There are a number of technological challenges,

which are being solved by various parts of the community using existing interferometers, including large-area coverage, multiple beams and large-scale data processing and correlation (e.g. LOFAR), transmission by long distances along optical fibre (e.g. e-MERLIN, EVLA), the possibility of high-resolution imaging using very long baselines and real-time correlation (e.g. eVLBI) and the problems of high frequencies (e.g. ALMA).

Although the SKA is some years in the future, the scientific potential is huge. For example, it will be able to see the faint emission of neutral hydrogen at cosmologically significant distances, give precision tests of cosmology and the star formation history of the universe by galaxy counts, resolve stellar disks and protoplanetary systems, search for extraterrestrial intelligence, and provide very sensitive tests of general relativity using pulsar studies.

Acknowledgement. I thank the organisers of the workshop for excellent organisation and hospitality. It was supported by the Jetset EU research training network programme MRTN-CT-2004-005592. Work at Jodrell Bank on interferometry and gravitational lenses is supported by the EU network MRTN-CT-2004-505183 “ANGLES”. I thank the Kavli Institute for Theoretical Physics for hospitality during the writing of this review, and Sir Francis Graham-Smith for a careful reading of the manuscript.

References

1. Adgie R.L., Gent H., Slee O.B., Frost A.B., Palmer H.P., Rowson B., 1965, Nature, 208, 275.
2. Allen R.J., Hamaker, J.P., Wellington, K.J., 1974, A&A 31, 17
3. Born M., Wolf E., Principles of Optics, Cambridge University Press. ISBN: 0-521-63921-2
4. Bryan R.K., Skilling J., 1980, MNRAS 191, 69
5. Burke B., Graham-Smith F., 2002, Cambridge University Press, ISBN 0-521-00517-5
6. Cornwell T., Wilkinson P.N., 1981, MNRAS 196, 1067
7. Davies J.G., Anderson B.A., Morison I., 1980, Nature 288, 64
8. Hogbom J.A. 1974, A&AS, 15, 417
9. Linfield R.P. et al., 1989, ApJ 336, 1105
10. Monnier J. 2003, Rep. Prog. Phys. 66, 789
11. Röttgering H, 2005, IAU Coll. 199, Shanghai, China, ed. Williams P. et al., Cambridge University Press, p. 281
12. Rowson B. 1963, MNRAS 125, 177
13. Ryle M., 1962, Nature, 194, 517
14. Ryle M., 1972, Nature, 239, 435
15. Swarup G., Ananthakrishnan S., Kapahi V.K., Rao A.P., Subrahmanya C.R., Kulkarni V.K., 1991, Current Science, 60, 2
16. Taylor G.B., Carilli C., Perley R.A. (eds) 1998, Astronomical Society of the Pacific Conf. Ser, vol. 180
17. Thompson A.R., Clark B.G., Wade C.M., Napier J.R., 1980, ApJS 44, 151

18. Thompson A.R., Moran J.M., Swenson G.W. 2001, ISBN 0-471-25492-4, Wiley.
19. Wilkinson P.N., 1991, In "Radio interferometry: Theory, techniques, and applications: Proceedings of the 131st IAU Colloquium, Socorro, Astronomical Society of the Pacific, 1991, p. 428.
20. Young J.S. et al. 2004, Proc National Astronomy Meeting, Milton Keynes, available from www.mrao.cam.ac.uk/telescopes/coast.