

Chapter 11

Further Imaging Techniques

This chapter is concerned with techniques of processing that are largely nonlinear and include deconvolution, that is removing, to the extent possible, the limitations of the visibility measurements. There are two principal deficiencies in the visibility data that limit the accuracy of synthesis images. These are (1) the limited distribution of spatial frequencies in u and v and (2) errors in the visibility measurements. The limited spatial frequency coverage can be improved by deconvolution processes such as CLEAN that allow the unmeasured visibility to take nonzero values within some general constraints on the image. Calibration can be improved by adaptive techniques in which the antenna gains, as well as the required image, are derived from the visibility data. Wide-field imaging, multifrequency imaging, and compressed sensing are also discussed.

11.1 The CLEAN Deconvolution Algorithm

One of the most successful deconvolution procedures is the algorithm CLEAN devised by Högbom (1974). This is basically a numerical deconvolving process usually applied in the image (l, m) domain. It has become an essential tool in producing images from incomplete (u, v) data sets. The procedure is to break down the intensity distribution into point-source responses that correspond to the original imaging process, and then replace each one with the corresponding response to a beam that is free of sidelobes. CLEAN can be thought of as a type of compressed sensing (see Sect. 11.8.6).

11.1.1 CLEAN Algorithm

The principal steps are as follows.

1. Compute the image and the response to a point source by Fourier transformation of the visibility and the weighted transfer function. These functions, the synthesized intensity and the synthesized beam, are often referred to as the “dirty image” and the “dirty beam,” respectively. The spacing of the sample points in the (l, m) plane should not exceed about one-third of the synthesized beamwidth.
2. Find the highest intensity point in the image and subtract the response to a point source, i.e., the dirty beam, including the full sidelobe pattern, centered on that position. The peak amplitude of the subtracted point-source response is equal to γ times the corresponding image amplitude. γ is called the loop gain, by analogy with negative feedback in electrical systems, and commonly has a value of a few tenths. Record the position and amplitude of the component removed by inserting a delta-function component into a model that will become the cleaned image.
3. Return to step 2 and repeat the procedure iteratively until all significant source structure has been removed from the image. There are several possible indicators of this condition. For example, one can compare the highest peak with the rms level of the residual intensity, look for the first time that the rms level fails to decrease when a subtraction is made, or note when significant numbers of negative components start to be removed.
4. Convolve the delta functions in the cleaned model with a clean-beam response, that is, replace each delta function with a clean-beam function of corresponding amplitude. The clean beam is often chosen to be a Gaussian with a half-amplitude width equal to that of the original synthesized (dirty) beam, or some similar function that is free from negative values.
5. Add the residuals (the residual intensity from step 3) into the clean-beam image, which is then the output of the process. (When the residuals are added, the Fourier transform of the image is equal to the measured visibilities.)

It is assumed that each dirty-beam response that is subtracted represents the response to a point source. As discussed in Sect. 4.4, the visibility function of a point source is a pair of real and imaginary sinusoidal corrugations that extend to infinity in the (u, v) plane. Any intensity feature for which the visibility function is the same within the (u, v) area sampled by the transfer function would produce a response in the image identical to the point-source response. Högbom (1974) has pointed out that much of the sky is a random distribution of point sources on an empty background, and CLEAN was initially developed for this situation. Nevertheless, experience shows that CLEAN also works on well-extended and complicated sources.

The result of the first three steps in the CLEAN procedure outlined above can be represented by a model intensity distribution that consists of a series of delta functions with magnitudes and positions representing the subtracted components. Since the modulus of the Fourier transform of each delta function extends uniformly

to infinity in the (u, v) plane, the visibility is extrapolated as required beyond the cutoff of the transfer function.

The delta-function components do not constitute a satisfactory model for astronomical purposes. Groups of delta functions with separations no greater than the beamwidth may actually represent extended structure. Convolution of the delta-function model by the clean beam, which occurs in step 4, removes the danger of overinterpretation. Thus, CLEAN performs, in effect, an interpolation in the (u, v) plane. Desirable characteristics of a clean beam are that it should be free from sidelobes, particularly negative ones, and that its Fourier transform should be constant inside the sampled region of the (u, v) plane and rapidly fall to a low level outside it. These characteristics are essentially incompatible since a sharp cutoff in the (u, v) plane results in oscillations in the (l, m) plane. The usual compromise is a Gaussian beam, which introduces a Gaussian taper in the (u, v) plane. Since this function tapers the measured data and the unmeasured data generated by CLEAN, the resulting intensity distribution no longer agrees with the measured visibility data. However, the absence of large, near-in sidelobes improves the dynamic range of the image, that is, it increases the range of intensity over which the structure of the image can reliably be measured.

As discussed in Chap. 10, we cannot directly divide out the weighted spatial transfer function on the right side of Eq. (10.4) because it is truncated to zero outside the areas of measurement. In CLEAN, this problem is solved by analyzing the measured visibility into sinusoidal visibility components and then removing the truncation so that they extend over the full (u, v) plane. Selecting the highest peak in the (l, m) plane is equivalent to selecting the largest complex sinusoid in the (u, v) plane.

At the point that the component subtraction is stopped, it is generally assumed that the residual intensity distribution consists mainly of the noise. Retaining the residual distribution within the image is, like the convolution with the clean beam, a nonideal procedure that is necessary to prevent misinterpretation of the final result. Without the residuals added in step 5, there would be an amplitude cutoff in the structure corresponding to the lowest subtracted component. Also, the presence of the background fluctuations provides an indication of the level of uncertainty in the intensity values. An example of the effect of processing with the CLEAN algorithm is shown in Fig. 11.1.

11.1.2 Implementation and Performance of the CLEAN Algorithm

As a procedure for removing sidelobe responses, CLEAN is easy to understand. Being highly nonlinear, however, CLEAN does not yield readily to a complete mathematical analysis. Some conclusions have been derived by Schwarz (1978, 1979), who has shown that conditions for convergence of CLEAN are that the synthesized beam must be symmetrical and its Fourier transform, that is, the

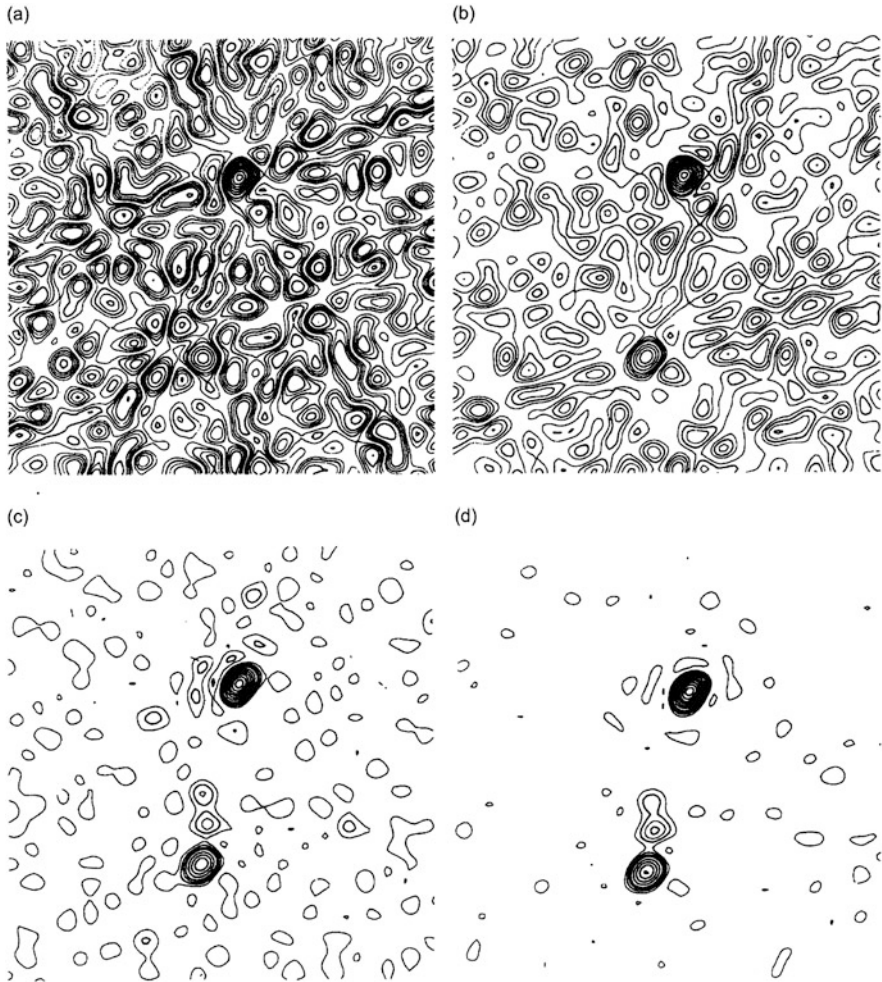


Fig. 11.1 Illustration of the CLEAN procedure using observations of 3C224.1 at 2695 MHz made with the interferometer at Green Bank, and rather sparse (u, v) coverage. (a) The synthesized “dirty” image; (b) the image after one iteration with the loop gain $\gamma = 1$; (c) after two iterations; (d) after six iterations. The components removed were restored with a clean beam in all cases. The contour levels are 5, 10, 15, 20, 30, etc., percent of the maximum value. From J. A. Högbom (1974), reproduced with permission. © ESO.

weighted transfer function, must be nonnegative. These conditions are fulfilled in the usual synthesis procedure. Schwarz’s analysis also indicates that if the number of delta-function components in the CLEAN model does not exceed the number of independent visibility data, CLEAN converges to a solution that is the least-mean-squares fit of the Fourier transforms of the delta-function components to the measured visibility. In enumerating the visibility data, either the real and imaginary

parts or the conjugate values (but not both) are counted independently. In images made using the fast Fourier transform (FFT) algorithm, there are equal numbers of grid points in the (u, v) and (l, m) planes, but not all (u, v) grid points contain visibility measurements. To maintain the condition for convergence, it is a common procedure to apply CLEAN only within a limited area, or “window,” of the original image.

In order to clean an image of a given dimension, it is necessary to have a dirty beam pattern of twice the image dimensions so that a point source can be subtracted from any location in the image. However, it is often convenient for the image and beam to be the same size. In that case, only the central quarter of the image can be properly processed. Thus, it is commonly recommended that the image obtained from the initial Fourier transform should have twice the dimensions required for the final image. As mentioned above, the use of such a window also helps to ensure that the number of components removed does not exceed the number of visibility data and, in the absence of noise, allows the residuals within the window area to approach zero.

Several arbitrary choices influence the result of the CLEAN process. These include the parameter γ , the window area, and the criterion for termination. Note that a point-source component in the image can be removed in one step of CLEAN only if it is centered on an image cell. This is an important reason for choosing $\gamma \ll 1$. A value between 0.1 and 0.5 is usually assigned to γ , and it is a matter of general experience that CLEAN responds better to extended structure if the loop gain is in the lower part of this range. The computation time for CLEAN increases rapidly as γ is decreased, because of the increasing number of subtraction cycles required. If the signal-to-noise ratio is \mathcal{R}_{sn} , then the number of cycles required for one point source is $-\log \mathcal{R}_{\text{sn}} / \log(1 - \gamma)$. Thus, for example, with $\mathcal{R}_{\text{sn}} = 100$ and $\gamma = 0.2$, a point source requires 21 cycles.

A well-known problem of CLEAN is the generation of spurious structure in the form of spots or ridges as modulation on broad features. A heuristic explanation of this effect is given by Clark (1982). The algorithm locates the maximum in the broad feature and removes a point-source component, as shown in Fig. 11.2. The negative sidelobes of the beam add new maxima, which are selected in subsequent cycles, and thus, there is a tendency for the component subtraction points to be located at intervals equal to the spacing of the first sidelobe of the synthesized (dirty) beam. The resulting image contains a lumpy artifact introduced by CLEAN, but the image is consistent with the measured visibility data. Cornwell (1983) introduced a modification of the CLEAN algorithm that is intended to reduce this unwanted modulation. The original CLEAN algorithm minimizes

$$\sum_k w_k |\mathcal{V}_k^{\text{meas}} - \mathcal{V}_k^{\text{model}}|^2, \quad (11.1)$$

where $\mathcal{V}_k^{\text{meas}}$ is the measured visibility at (u_k, v_k) , w_k is the applied weighting, and $\mathcal{V}_k^{\text{model}}$ is the corresponding visibility of the CLEAN-derived model. The summation is taken over the points with nonzero data in the input transformation for the dirty



Fig. 11.2 Subtraction of the point-source response (broken line) at the maximum of a broad feature, as in the process CLEAN. Adapted from Clark (1982).

image. Cornwell's algorithm minimizes

$$\sum_k w_k |\mathcal{V}_k^{\text{meas}} - \mathcal{V}_k^{\text{model}}|^2 - \kappa s, \quad (11.2)$$

where s is a measure of smoothness, and κ is an adjustable parameter. Cornwell found that the mean-squared intensity of the model, taken with a negative sign, is an effective implementation of s .

The effects of visibility tapering appear in both the original image and the beam, and thus the magnitudes and positions of the components subtracted in the CLEAN process should be largely independent of the taper. However, since tapering reduces the resolution, it is a common practice to use uniform visibility weighting for images that are processed using CLEAN. Alternately, in difficult cases such as those involving extended smooth structure, reduction of sidelobes by tapering may improve the performance of CLEAN.

An important reduction in the computation required for CLEAN was introduced by Clark (1980). This is based on subtraction of the point-source responses in the (u, v) plane and using the FFT for moving data between the (u, v) and (l, m) domains. The procedure consists of minor and major cycles. A series of minor cycles is used to locate the components to be removed by performing approximate subtractions using only a small patch of the synthesized dirty beam that includes the main beam and the major sidelobes. Then in a major cycle, the identified point-source responses are subtracted, without approximation, in the (u, v) plane. That is, the convolution of the delta functions with the dirty beam is performed by multiplying their Fourier transforms. The series of minor and major cycles is then repeated until the required stop condition is reached. Clark devised this technique for use with data from the VLA and found that it reduced the computation by a factor of two to ten compared with the original CLEAN algorithm.

Other variations on the CLEAN process have been devised; one of the more widely used is the Cotton–Schwab algorithm [Schwab (1984); see Sect. IV], which is a variation of the Clark algorithm. The subtractions in the major cycle are performed on the ungridded visibility data, which eliminates aliasing at this point. The algorithm is also designed to permit processing of adjacent fields, which are treated separately in the minor cycles but in the major cycles, components are jointly removed from all fields.

To summarize the characteristics of CLEAN, we note that it is simple to understand from a qualitative viewpoint and straightforward to implement and that

its usefulness is well proven. On the other hand, a full analysis of its response is difficult. The response of CLEAN is not unique, and it can produce spurious artifacts. It is sometimes used in conjunction with model-fitting techniques; for example, a disk model can be removed from the image of a planet and the residual intensity processed by CLEAN. A more stable and efficient version of CLEAN called multiscale CLEAN has been developed for extended objects (Wakker and Schwarz 1988; Cornwell 2008). The basic idea is that broad emission components are identified first and removed. More sophisticated methods are being developed to handle extended emission [e.g., Junklewitz et al. (2016)]. CLEAN is also used as part of more complex image construction techniques. For more details, including hints on usage, see Cornwell et al. (1999), and for extended objects, Cornwell (2008).

11.2 Maximum Entropy Method

11.2.1 MEM Algorithm

An important class of image-restoration algorithms operates to produce an image that agrees with the measured visibility to within the noise level, while constraining the result to maximize some measure of image quality. Of these, the maximum entropy method (MEM) has received particular attention in radio astronomy. If $I'(l, m)$ is the intensity distribution derived by MEM, a function $F(I')$ is defined, which is referred to as the entropy of the distribution. $F(I')$ is determined entirely by the distribution of I' as a function of solid angle and takes no account of structural forms within the image. In constructing the image, $F(I')$ is maximized within the constraint that the Fourier transform of I' should fit the observed visibility values.

In astronomical image formation, an early application of MEM is that of Frieden (1972) to optical images. In radio astronomy, the earliest discussions are by Ponsonby (1973) and Ables (1974). The aim of the technique, as described by Ables, is to obtain an intensity distribution consistent with all relevant data but minimally committal with regard to missing data. Thus, $F(I')$ must be chosen so that maximization introduces legitimate a priori information but allows the visibility in the unmeasured areas to assume values that minimize the detail introduced.

Several forms of $F(I')$ have been used, which include the following:

$$F_1 = - \sum_i \frac{I'_i}{I'_s} \log \left(\frac{I'_i}{I'_s} \right) \quad (11.3a)$$

$$F_2 = - \sum_i \log I'_i \quad (11.3b)$$

$$F_3 = - \sum_i I'_i \ln \left(\frac{I'_i}{M_i} \right), \quad (11.3c)$$

where $I'_i = I'(l_i, m_i)$, $I'_s = \sum_i I'_i$, M_i represents an a priori model, and the sums are taken over all pixels, I_i , in the image. F_3 can be described as relative entropy, since the intensity values are specified relative to a model.

A number of papers discuss the derivation of the expressions for entropy from theoretical and philosophical considerations. Bayesian statistics are invoked: see Jaynes (1968, 1982). Gull and Daniell (1979) consider the distributions of intensity quanta scattered randomly on the sky, and they derive the form F_1 , which is also used by Frieden (1972). The entropy form F_2 is obtained by Ables (1974) and Wernecke and D'Addario (1977). Other investigators take a pragmatic approach to MEM (Högbom 1979, Subrahmanya 1979, Nityananda and Narayan 1982). They view the method as an effective algorithm, even though there may be no underlying physical or information-theoretical basis for the choice of constraints. Högbom (1979) points out that both F_1 and F_2 contain the required mathematical characteristics: the first derivatives tend to infinity as I' approaches zero, so maximizing F_1 or F_2 produces positivity in the image. The second derivatives are everywhere negative, which favors uniformity in the intensity. Narayan and Nityananda (1984) consider a general class of functions F that have the properties $d^2F/dI'^2 < 0$ and $d^3F/dI'^3 > 0$. F_1 and F_2 , discussed above, are members of this class.

In the maximization of the entropy expression $F(I')$, the constraint that the resulting intensity model should be consistent with the measured visibility data is implemented through a χ^2 statistic. Here, χ^2 is a measure of the mean-squared difference between the measured visibility values, $\mathcal{V}_k^{\text{meas}} = \mathcal{V}(u_k, v_k)$, and the corresponding values for the model $\mathcal{V}_k^{\text{model}}$:

$$\chi^2 = \sum_k \frac{|\mathcal{V}_k^{\text{meas}} - \mathcal{V}_k^{\text{model}}|^2}{\sigma_k^2}, \quad (11.4)$$

where σ_k^2 is the variance of the noise in $\mathcal{V}_k^{\text{meas}}$, and the summation is taken over the visibility data set. Obtaining a solution involves an iterative procedure; see Wernecke and D'Addario (1977), Wernecke (1977), Gull and Daniell (1978), Skilling and Bryan (1984), and a review by Narayan and Nityananda (1984). As an example, Cornwell and Evans (1985) maximize a parameter J given by

$$J = F_3 - \alpha\chi^2 - \beta S_{\text{model}}, \quad (11.5)$$

where F_3 is defined in Eq. (11.3c). S_{model} is the total flux density of the model and is included because in order for the process to converge to a satisfactory result, it was found necessary to include a constraint that the total flux density of the model be equal to the measured flux density. Lagrange multipliers α and β are included, the values of which are adjusted as the model fitting proceeds so that χ^2 and S_{model} are equal to the expected values. Through the use of F_3 , a priori information can be introduced into the final image. The various algorithms that have been developed for implementing MEM generally use the gradients of the entropy and of χ^2 to determine the adjustment of the model in each iteration cycle.

A feature of images derived by MEM is that the point-source response varies with position, so the angular resolution is not constant over the image. Comparison of maximum entropy images with those obtained using direct Fourier transformation often shows higher angular resolution in the former. The extrapolation of the visibility values can provide some increase in resolution over more conventional imaging techniques.

11.2.2 Comparison of CLEAN and MEM

CLEAN is defined in terms of a procedure, so the implementation is straightforward, but because of the nonlinearity in the processing, a noise analysis of the result is very difficult. In contrast, MEM is defined in terms of an image that fits the data to within the noise and is also constrained to maximize some parameter of the image. The noise in MEM is taken into account through the χ^2 statistic, and the resulting effect on the noise is more easily analyzed for MEM; see, for example, Bryan and Skilling (1980). Some further points of comparison are as follows:

- Implementation of MEM requires an initial source model, which is not necessary in CLEAN.
- CLEAN is usually faster than MEM for small images, but MEM is faster for very large images. Cornwell et al. (1999) give the break-even point as about 10^6 pixels.
- CLEAN images tend to show a small-scale roughness, attributable to the basic approach of CLEAN, which models all images as ensembles of point sources. In MEM, the constraint in the solution emphasizes smoothness in the image.
- Broad, smooth features are better deconvolved using MEM, since CLEAN may introduce stripes and other erroneous detail. MEM does not perform well on point sources, particularly if they are superimposed on a smooth background that prevents negative sidelobes from appearing as negative intensity in the dirty image.

To illustrate the characteristics of the CLEAN and MEM procedures, Fig. 11.3 shows examples of processing of a model jet structure from Cornwell (1995) and Cornwell et al. (1999), using model calculations by Briggs. The jet model is based on similar structure in M87 and is virtually identical to the contour levels shown in part (e). The left end of the jet is a point source smoothed to the resolution of the simulated observation. Visibility values for the model corresponding to the (u, v) coverage of the VLBA (Napier et al. 1994) were calculated for a frequency of 1.66 GHz and a declination of 50° with essentially full tracking range. Thermal noise was added, but the calibration was assumed to be fully accurate. Fourier transformation of the visibility data and the spatial transfer function provided the dirty image and dirty beam. The image shows the basic structure, but fine details are swamped by sidelobes. Parts (a) to (c) of Fig. 11.3 show the effects of processing by CLEAN. In the CLEAN deconvolution, 20,000 components were subtracted with a

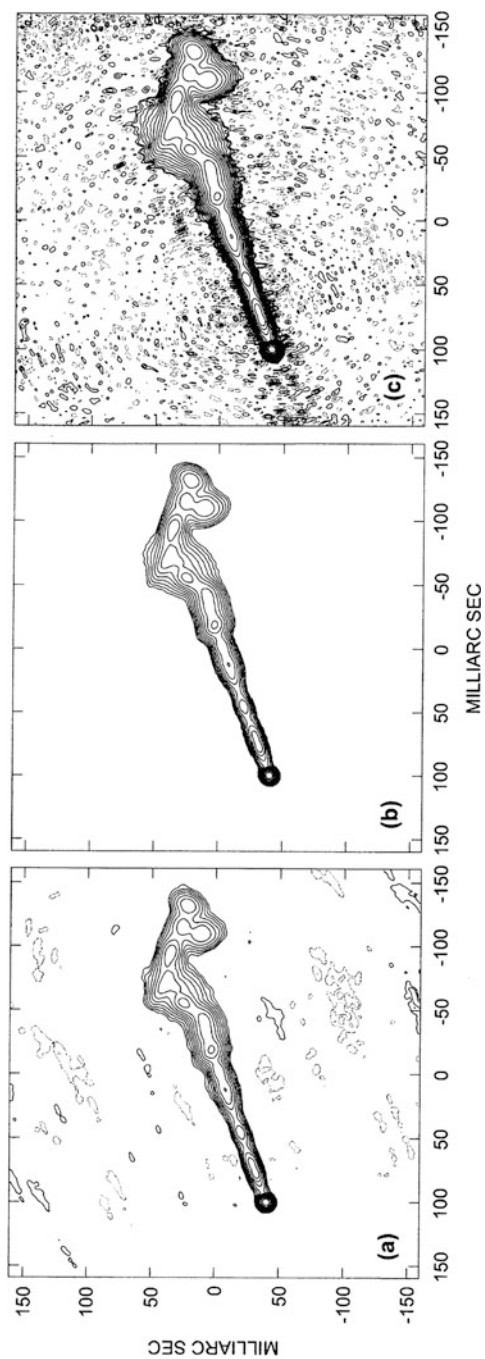


Fig. 11.3 Examples of deconvolution procedures applied to a model jet structure that includes a point source at the left end. Part (a) shows the result of application of CLEAN to the whole image, and (b) the result when components are taken only within a tight support region surrounding the source. Note the improvement obtained in (b). The contours approximately indicate the intensity increasing in powers of two from a low value of 0.05%. Part (c) shows the same image as (b) but with contours starting a factor of ten lower in intensity, and the roughness characteristic of CLEAN is visible in the low-level contours.

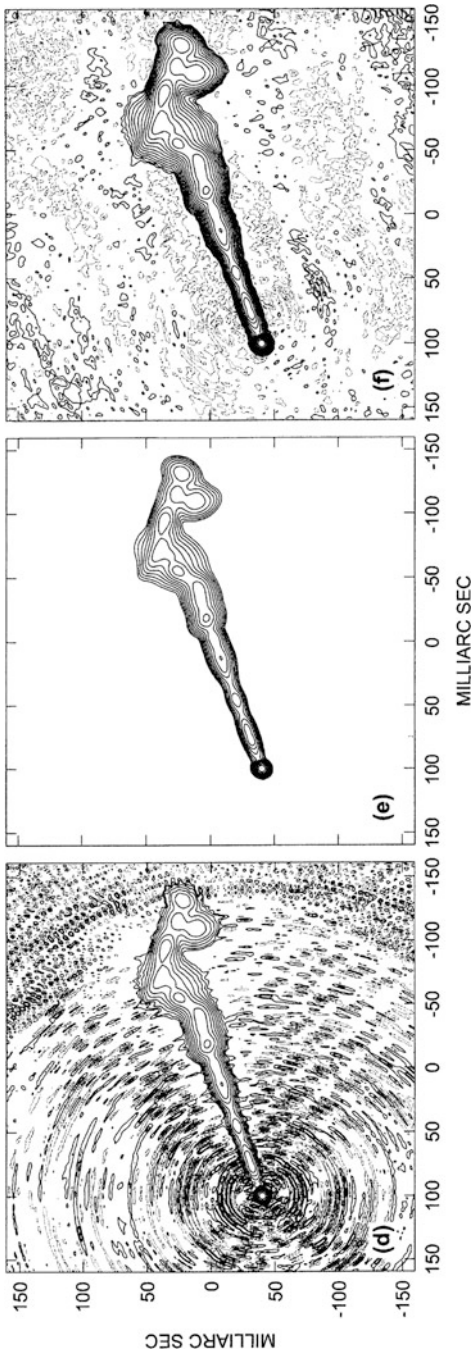


Fig. 11.3 Part (d) shows the result of MEM deconvolution using the same constraint region as in (b) and 80 iterations. The circular artifacts, centered on the point source, illustrate the well-known inability of MEM to handle sharp features well. In part (e), the point source was subtracted, using the CLEAN response to the feature, and then the MEM deconvolution was performed with the same constraint region as in (d). The source was then replaced. Part (f) shows the same response as (e) with the lowest contours at the same level as part (e). Note that the low-level contours are smoother in the MEM image than in the CLEAN one. The images in (c) and (f) show comparable fidelity to the model. All six parts are from Cornwell (1995), courtesy of and © the Astronomical Society of the Pacific.

loop gain of 0.1. Part (a) shows the result of application of CLEAN to the whole image, and part (b) shows the result when components are taken only within a tight support region surrounding the source (the technique sometimes referred to as use of a box or window). Note the improvement obtained in (b), which is a result of adding the information that there is no emission outside the box region. The contours approximately indicate the intensity increasing in powers of two from a low value of 0.05%. Part (c) shows the same image as panel (b) but with contours starting a factor of ten lower in intensity. The roughness visible in the low-level contours is characteristic of CLEAN, in which each component is treated independently and there is no mechanism to relate the result for any one component to those for its neighbors, unlike the case of MEM, in which a smoothness constraint is introduced. Parts (d) to (f) result from MEM processing. Part (d) shows the result of MEM deconvolution using the same constraint region as in panel (b) and 80 iterations. The circular pattern of the background artifacts, centered on the point source, clearly shows that MEM does not handle such a feature well. In part (e), the point source was subtracted, using the CLEAN response to the feature, and then the MEM deconvolution performed with the same constraint region as in (d). The source was then replaced. Part (f) shows the same response as (e) with the lowest contours at the same level as panel (c). The low-level contours show the structure contributed by the observation and processing. The contours are smoother in the MEM image than in the CLEAN one. The images in (c) and (f) have comparable *fidelity*, that is, accuracy of reproduction of the initial model. Combinations of procedures, such as the use of CLEAN to remove point-source responses from an image and then the use of MEM to process the broader background features can sometimes be used to advantage in complex images.

11.2.3 Further Deconvolution Procedures

Briggs (1995) has applied a nonnegative, least-squares (NNLS) algorithm for deconvolution. The NNLS algorithm was developed by Lawson and Hanson (1974) and provides a solution to a matrix equation of the form $\mathbf{AX} = \mathbf{B}$, where, in the radio astronomy application, \mathbf{A} represents the dirty beam and \mathbf{B} the dirty image. The algorithm provides a least-mean-squares solution for the intensity \mathbf{X} that is constrained to contain no negative values. However, unlike the case for MEM, there is no smoothness criterion involved. The NNLS solution requires more computer capacity than CLEAN or MEM solutions, but Briggs's investigation indicated that it is capable of superior performance, particularly in cases of compact objects of width only a few synthesized beamwidths. NNLS was found to reduce the residuals to a level close to the system noise in the observations. In certain cases, it was found to work more effectively than CLEAN in hybrid imaging and self-calibration procedures (discussed in Sect. 11.3) and to allow higher dynamic range to be achieved. In MEM, the residuals may not be entirely random but may be correlated in the image plane, and this effect can introduce bias in the (u, v) data that limits the

achievable dynamic range. CLEAN appears to behave somewhat similarly unless it is allowed to run long enough to work down into the noise. Some further discussion can be found in Briggs (1995) and Cornwell et al. (1999).

11.3 Adaptive Calibration and Imaging

Calibration of the visibility amplitude is usually accurate to a few percent or better, but phase errors expressed as a fraction of a radian may be much larger, sometimes as a result of variations in the ionosphere or troposphere. Nevertheless, the relative values of the uncalibrated visibility measured simultaneously on a number of baselines contain information about the intensity distribution that can be extracted through the closure relationships described in Chap. 10, Eqs. (10.34) and (10.44). Following Schwab (1980), we use the term *adaptive calibration* for both the hybrid imaging and self-calibration techniques that make use of this information. Imaging with amplitude data only has also been investigated and is briefly described.

11.3.1 Hybrid Imaging

The rekindling of interest in closure techniques in the 1970s began with their rediscovery by Rogers et al. (1974), who used closure phases to derive model parameters for VLBI data. Fort and Yee (1976) and several later groups incorporated closure data into iterative imaging techniques, of which that by Readhead et al. (1980) is as follows:

1. Obtain an initial trial image based on inspection of visibility amplitudes and any a priori data such as an image at a different wavelength or epoch. If the trial image is inaccurate, the convergence will be slow, but if necessary, an arbitrary trial image such as a single point source will often suffice.
2. For each visibility integration period, determine a complete set of independent amplitude and/or phase closure equations. For each such set, compute a sufficient number of visibility values from the model such that when added to the closure relationships, the total number of independent equations is equal to the number of antenna spacings.
3. Solve for the complex visibility corresponding to each antenna spacing and make an image from the visibility data by Fourier transformation.
4. Process the image from step 3 using CLEAN but omitting the residuals.
5. Apply constraints for positivity and confinement (delete components having negative intensity or lying outside the area that is judged to contain the source).
6. Test for convergence and return to step 2 as necessary, using the image from step 5 as the new model.

Note that the solution improves with iteration because of the constraints of confinement and positivity introduced in step 5. These nonlinear processes can be envisioned as spreading the errors in the model-derived visibility values throughout

the visibility data, so that they are diluted when combined with the observed values in the next iterative cycle.

In the process described, and most variants of it, the image is formed by using some data from the model and some from direct measurements, and following Baldwin and Warner (1978), the term *hybrid* imaging (or mapping) is sometimes used as a generic description. With the use of phase closure, there is no absolute position measurement, but there is no ambiguity with respect to the position angle of the image. With the use of amplitude closure, only relative levels of intensity are determined, but it is usually not difficult to calibrate enough of the data to establish an intensity scale. In many cases, the amplitude data are sufficiently accurate as observed, and only the phase closure relationships need be used; Readhead and Wilkinson (1978) have described a version of the above program using phase closure only. Other versions of this technique, which differ mainly in detail of implementation from that described, have been developed by Cotton (1979) and Rogers (1980). If there is some redundancy in the baselines, the number of free parameters is reduced, which can be advantageous, as discussed by Rogers.

The number of antennas, n_a , is obviously an important factor in imaging procedures that make use of the closure relationships, since it affects the efficiency with which the data are used. We can quantify this efficiency by considering the number of closure data as a fraction of the number of data that would be available if full calibration were possible, as a function of n_a . The numbers of independent closure data are given by Eqs. (10.42) and (10.45). The number of data with full calibration is equal to the number of baselines, which, if we assume there is no redundancy, is $\frac{1}{2}n_a(n_a - 1)$. For the phase data, the fraction is

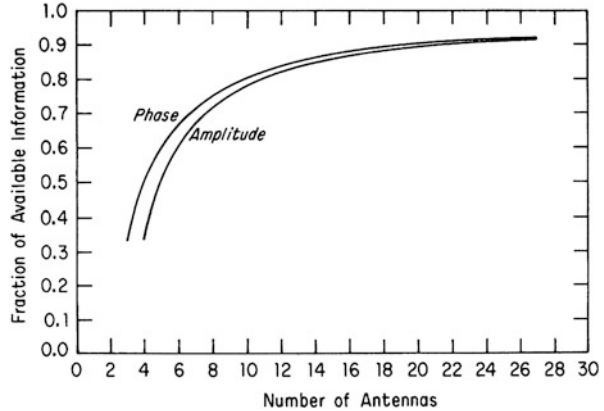
$$\frac{\frac{1}{2}(n_a - 1)(n_a - 2)}{\frac{1}{2}n_a(n_a - 1)} = \frac{n_a - 2}{n_a} . \quad (11.6)$$

For the amplitude data, the fraction is

$$\frac{\frac{1}{2}n_a(n_a - 3)}{\frac{1}{2}n_a(n_a - 1)} = \frac{n_a - 3}{n_a - 1} . \quad (11.7)$$

These fractions are also equal to the ratios of observed data to observed plus model-derived data in each iteration of the hybrid imaging procedure. Equations (11.6) and (11.7) are plotted in Fig. 11.4. For $n_a = 4$, the closure relationships yield only 50% of the possible phase data and only 33% of the amplitude data. For $n_a = 10$, however, the corresponding figures are 80% and 78%. Thus, in any array in which the atmosphere or instrumental effects may limit the accuracy of calibration by a reference source, it is desirable that the number of antennas should be at least ten and preferably more. The number of iterations required to obtain a solution with the hybrid technique depends on the complexity of the source, the number of antennas, the accuracy of the initial model, and other factors including details of the algorithm used.

Fig. 11.4 Visibility data that can be obtained through adaptive calibration techniques expressed as a fraction of those available from a fully calibrated array. The curves correspond to Eqs. (11.6) and (11.7).



11.3.2 Self-Calibration

Hybrid imaging has largely been superseded by a more general approach called *self-calibration*. Here, the complex antenna gains are regarded as free parameters to be explicitly derived together with the intensity. In certain cases, the process is easily explained. For example, in imaging an extended source containing a compact component (as in many radio galaxies), the broad structure is resolved with the longer antenna spacings, leaving only the compact source. This can be used as a calibrator to provide the relative phases of the long-spacing antenna pairs, but not the absolute phase since the position is not known. Then, if there is a sufficient number of long spacings in the array, the relative gain factors of the antennas can be obtained using long spacings only. Such a special intensity distribution, however, is not essential to the method, and with an iterative technique, it is possible to use almost any source as its own calibrator. Programs of this type were developed by Schwab (1980) and by Cornwell and Wilkinson (1981). Reviews of the techniques are given by Pearson and Readhead (1984) and Cornwell (1989).

The procedure in self-calibration is to use a least-mean-squares method to minimize the square of the modulus of the difference between the observed visibilities, $\mathcal{V}_{mn}^{\text{meas}}$, and the corresponding values for the derived model, $\mathcal{V}_{mn}^{\text{model}}$. The expression that is minimized is

$$\sum_{\text{time}} \sum_{m < n} w_{mn} |\mathcal{V}_{mn}^{\text{meas}} - g_m g_n^* \mathcal{V}_{mn}^{\text{model}}|^2, \tag{11.8}$$

where the weighting coefficient w_{mn} is usually chosen to be inversely proportional to the variance of $\mathcal{V}_{mn}^{\text{meas}}$, and the quantities shown are all functions of time within the observing period. Expression (11.8) can be written

$$\sum_{\text{time}} \sum_{m < n} w_{mn} |\mathcal{V}_{mn}^{\text{model}}|^2 |X_{mn} - g_m g_n^*|^2, \tag{11.9}$$

where

$$X_{mn} = \frac{\mathcal{V}_{mn}^{\text{meas}}}{\mathcal{V}_{mn}^{\text{model}}}. \quad (11.10)$$

If the model is accurate, the ratio X_{mn} of the uncalibrated observed visibility to the visibility predicted by the model is independent of u and v but proportional to the antenna gains. Thus, the values of X_{mn} simulate the response to a calibrator and enable the gains to be determined. However, since the initial model is only approximate, the desired result must be approached by iteration.

The self-calibration procedure is:

1. Make an initial image as for hybrid imaging.
2. Compute the X_{mn} factors for each visibility integration period within the observation.
3. Determine the antenna gain factors for each integration period.
4. Use the gains to calibrate the observed visibility values and make an image.
5. Use CLEAN and select components to provide positivity and confinement of the image; Cornwell (1982) recommends omitting all features for which $|I(l, m)|$ is less than that for the most negative feature.
6. Test for convergence and return to step 2 as necessary.

The numbers of independent data used in the procedure above are, as in the case of hybrid imaging, equal to the numbers of independent closure relationships given in Eqs. (10.45) and (10.36), that is, $\frac{1}{2}n_a(n_a - 3)$ for amplitude and $\frac{1}{2}(n_a - 1)(n_a - 2)$ for phase. The two procedures, hybrid imaging and self-calibration, are basically equivalent but differ in details of approach and implementation. The efficiency as a function of the number of antennas (Fig. 11.4) applies to both. Examples of the performance of the self-calibration technique are shown in Figs. 11.5 and 11.6.

Treating the gain factors, which are the fundamental unknown quantities, as free parameters as in self-calibration is a rather more direct approach than that of hybrid imaging. A global estimate of the instrumental factors is obtained using the entire data set. Cornwell (1982) points out that it is easier to deal correctly with the noise when considering complex visibility as a vector quantity, as in self-calibration, than when considering amplitude and phase separately, as in hybrid imaging. The noise combines additively in the vector components, resulting in a Gaussian distribution, whereas in the amplitude and phase, the more complicated Rice distributions of Eqs. (6.63) result. Cornwell and Wilkinson (1981) have developed a form of adaptive calibration that takes account of the different probability distributions of the amplitude and phase fluctuations, including system noise, for the different antennas. It has been used with the MERLIN array of Jodrell Bank, UK, which incorporates antennas of different sizes and designs (Thomasson 1986). The probability distributions of the antenna-associated errors are legitimate a priori information, which can be empirically determined for an array.

Experience shows that adaptive calibration techniques in many cases converge to a satisfactory result using only a single point source as a starting model, although

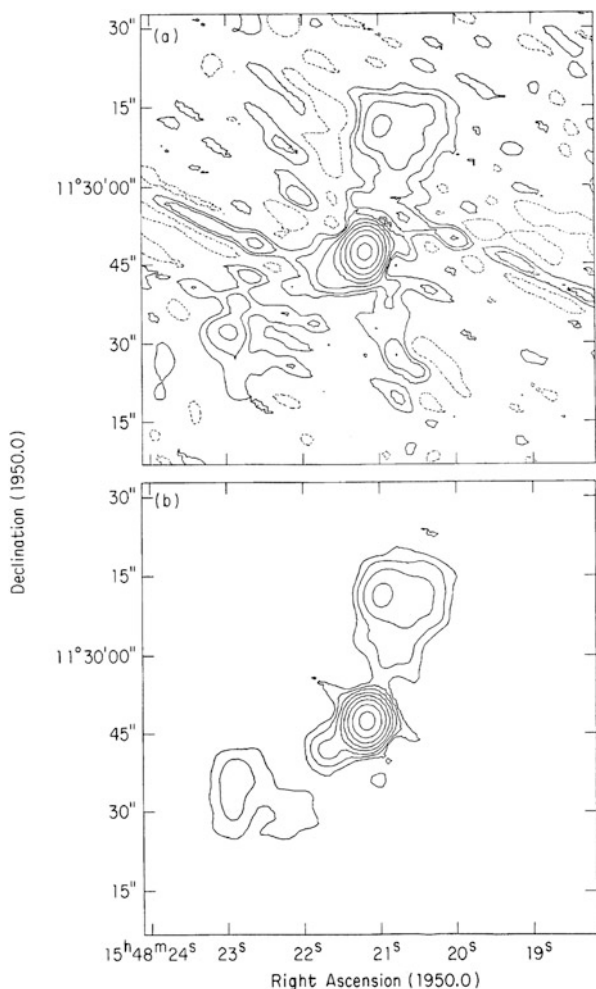


Fig. 11.5 Effect of self-calibration on a VLA radio image of the quasar 1548+115. (a) Image obtained by normal calibration techniques, which has spurious detail at the level of 1% of the peak intensity. (b) Image obtained by the self-calibration technique, in which the level of spurious detail is reduced below the 0.2% level. In both (a) and (b), the lowest contour level is 0.6%. © 1983 IEEE. Reprinted, with permission, from Napier et al. (1983).

inaccuracy in the initial model increases the number of iterative cycles required. A point source is a good model for the phase of a symmetrical intensity distribution but may be a poor model for the amplitude. It must also be remembered that the accuracy of the closure relationships depends on the accuracy of the matching of the frequency responses and polarization parameters from one antenna to another, as discussed in Sects. 7.3 and 7.4. In general, any effect that cannot be represented by a single gain factor for each antenna degrades the closure accuracy.

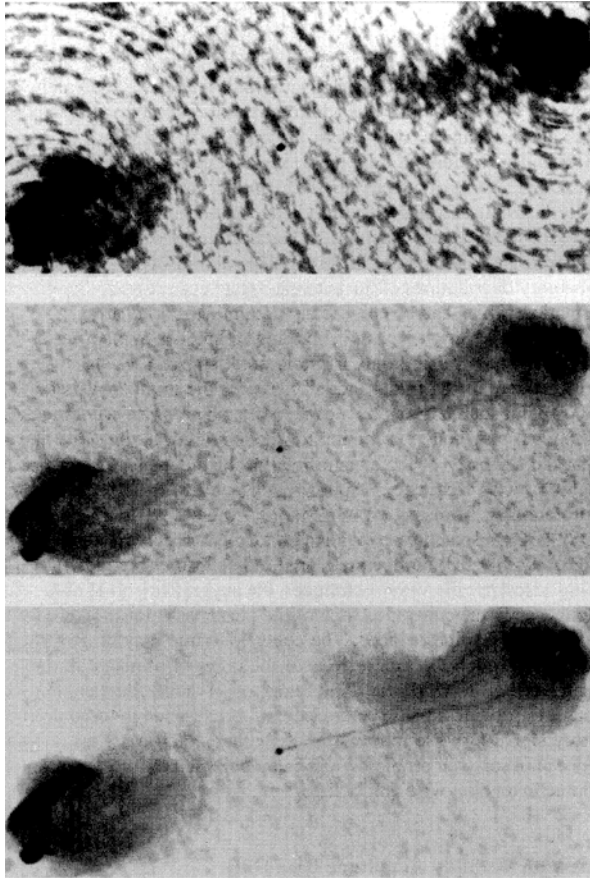


Fig. 11.6 Three stages in the reduction of the observation of Cygnus A shown in Fig. 1.18. The top image is the result of transformation of the calibrated visibility data using the FFT algorithm. The calibration source was approximately 3° from Cygnus A. The center image shows reduction using the MEM algorithm. This compensates principally for the undersampling in the spatial frequencies and thereby removes sidelobes from the synthesized beam. The result is similar to that obtainable using the CLEAN algorithm. The bottom image shows the effect of the self-calibration technique, in which the maximum entropy image is used as the initial model. The final step improves the dynamic range by a factor of 3. In observations in which the initial calibration is not as good as in this case, self-calibration usually provides a greater improvement. The long dimension of the field is $2.1'$ and contains approximately 1000 pixels. Courtesy R. A. Perley, J. W. Dreher, and J. J. Cowan. Reproduced by permission of and © NRAO/AUI.

In using adaptive calibration techniques, the integration period of the data must not be longer than the coherence time of the phase variations; otherwise, the visibility amplitude may be reduced. The coherence time may be governed by the atmosphere, for which the timescale is of the order of minutes (see Sect. 13.4). In order for the imaging procedure to work, the field under observation must contain structure fine enough to provide a phase reference and bright enough to be detected

with satisfactory signal-to-noise ratio within the coherence time. Thus, adaptive calibration does not solve all problems and cannot be used for the detection of a very weak source in an otherwise empty field.

11.3.3 Imaging with Visibility Amplitude Data Only

A number of early studies were made concerning the feasibility of producing images using only the amplitude values of the visibility. The Fourier transform of the squared modulus of the visibility is equal to the autocorrelation of the intensity distribution, $I \star \star I$:

$$|\mathcal{V}(u, v)|^2 = \mathcal{V}(u, v) \mathcal{V}^*(u, v) \longleftrightarrow I(l, m) \star \star I(l, m) . \quad (11.11)$$

The right side can also be written as a convolution: $I(l, m) * * I(-l, -m)$. The problem of imaging with $|\mathcal{V}|$ only is mainly one of interpreting an image of the autocorrelation of I . Without phase data, the position of the center of the field cannot be determined, and there is a 180° rotational ambiguity in the position angle of the image.

Examples of studies relevant to imaging without phase data are found in Bates (1969, 1984), Napier (1972), and Fienup (1978). Napier and Bates (1974) review some of the results. The positivity requirement is generally found to be insufficient to provide unique solutions for one-dimensional profiles, but for two-dimensional images, uniqueness is obtained in some cases (Bruck and Sodin 1979). Baldwin and Warner (1978, 1979) considered a two-dimensional distribution of sources, with some success in producing a source image from the autocorrelation function. Although these approaches showed some promise of providing useful interpretation of radio interferometer data, they have not been widely used. More importantly, the development of techniques that make use of closure relationships, such as hybrid imaging and self-calibration, has allowed visibility phases to contribute useful data even when not well calibrated.

11.4 Imaging with High Dynamic Range

The *dynamic range* of an image is usually defined as the ratio of the maximum intensity to the rms level at some part of the field where the background is mainly blank sky. This rms level is assumed to indicate the lowest measurable intensity. The term *image fidelity* is used to indicate the degree to which an image is an accurate representation of a source on the sky. Image fidelity is not directly measurable on an actual source, but simulation of an observation of a model source and reduction of the visibility data allow comparison of the resulting image with the model. This is a

way of investigating antenna configurations, processing methods, and other details. The requirements and techniques are discussed in detail by Perley (1989, 1999a).

High dynamic range requires high accuracy in calibration, removal of any erroneous data, and careful deconvolution. That is, it requires high accuracy in the visibility measurements and very good (u, v) coverage. A phase error $\Delta\phi$ can be regarded as introducing an erroneous component of relative amplitude, $\sin \Delta\phi$, into the visibility data, in phase quadrature to the true visibility. An amplitude error of $\varepsilon_a\%$ can be regarded as introducing an error component of relative amplitude $\varepsilon_a\%$ into the visibility. Thus, for example, a phase error of 10° introduces as large an error component as does an amplitude error of 17%. An amplitude error of 17% would be considered unusually large in most cases, except in conditions of strong atmospheric attenuation. However, a 10° phase error would be more commonly encountered, especially at frequencies in which ionospheric or tropospheric irregularities are important. A phase error $\Delta\phi$ (rad) in a correlator output introduces an error component of rms relative amplitude $\Delta\phi/\sqrt{2}$ in the resulting image. With similar errors in $n_a(n_a - 1)/2$ baselines, the dynamic range of a snapshot is limited to $\sim n_a/\Delta\phi$.

Use of self-calibration is an essential step in minimizing gain errors. However, after calibration of the antenna-based gain factors, there remain small baseline-based terms that can also be calibrated. These result from variations, from one antenna to another, in the frequency passband or the polarization, and similar effects. Note that in arrays with very high sensitivity at the longer wavelengths, the requirement to observe down to the limit set by system noise, in the presence of background sources, places a lower limit on the required dynamic range. A large number of array elements is helpful in distinguishing individual sources (Lonsdale et al. 2000). Braun (2013) describes a detailed analysis of dynamic range in synthesis imaging and gives the results of application of this analysis to several large arrays.

Obtaining the highest possible dynamic range requires attention to details that are specific to particular instruments. For the VLA, the following figures were quoted as rough guidelines for a good observation. Basic calibration results in dynamic range of order 1,000 : 1. After self-calibration, dynamic range up to $\sim 20,000$: 1 is possible. After careful correction of baseline-based errors, it may be a few times higher. If a spectral correlator is used, which avoids errors in quadrature networks and also relaxes the requirement for delay accuracy, a dynamic range of $\sim 200,000$: 1 is possible, with much care, assuming that the signal-to-noise ratio is adequate (Perley 1989).

11.5 Mosaicking

Mosaicking is a technique that allows imaging of an area of sky that is larger than the beam of the array elements. It becomes very important in the millimeter-wavelength range, where antenna beams are relatively narrow. Although radio

astronomy antennas for millimeter wavelengths are generally smaller in diameter than are antennas for centimeter wavelengths, their beamwidths are often narrower because the wavelengths are so much shorter. For example, the Atacama Large Millimeter/submillimeter Array (ALMA) can operate at frequencies up to 950 GHz, at which the beamwidth of the 12-m-diameter antennas could be as small as $\sim 6''$.

Consider imaging a square field whose sides are n times the width of the antenna primary beam. One can divide the required area into n^2 subfields, each the size of a beam, and produce a separate image for each such area. The n^2 beam-area images can then be fitted together like mosaic pieces to cover the full field desired. One would anticipate that some difficulty might occur in obtaining uniform sensitivity, particularly near the joints of the mosaic pieces, but clearly the idea is feasible. From the sampling theorem described in Sect. 5.2, the number of visibility sample points in u and v required in an image covering n^2 beam areas is n^2 times as many as would be required in an image that covers just one beam area. In mosaicking, the increased data are obtained by using n^2 different pointing directions of the antennas. As a result, the sampling of the visibility in u and v must be at an interval $1/n$ of that for a field equal to the beam size, and this interval is usually less than the diameter of the antenna aperture. However, it is possible to determine how the visibility varies on a scale less than the diameter of an antenna, as discussed below.

Figure 5.9 shows two antennas that are tracking the position of a source. The antenna spacing projected normal to the direction of the source is u , and the antenna diameter is d_λ , both quantities being measured in wavelengths. In the u direction, the interferometer responds to spatial frequencies from $(u - d_\lambda)$ to $(u + d_\lambda)$, since spacings within this range can be found within the antenna apertures. Measurement of the variation of the visibility over this range of baselines can provide the fine sampling required in mosaicking. The difference in path lengths from the source to the two antenna apertures is w wavelengths, and as the antennas track, the variation in w gives rise to fringes at the correlator output. Since the apertures of the antennas remain normal to the direction of the source, the path difference w , and its rate of change, are the same for any pair of points of which one is in each aperture plane, regardless of their spacing. Thus, because of the tracking motion, the signals received at any two such points produce a component of the correlator output with the same fringe frequency. Such components cannot, therefore, be separated by Fourier analysis, and information on the variation of the visibility within the spatial frequency range $(u - d_\lambda)$ to $(u + d_\lambda)$ is lost. However, in mosaicking, the antenna beams are scanned across the field, either by moving periodically between different pointing centers or by continuously scanning, for example, in a raster pattern. The scanning is in addition to the usual tracking motion to follow the source across the sky. In Fig. 5.9, it can be seen that if the antennas are suddenly turned through a small angle $\Delta\theta$, then the position of the point B is changed by $\Delta u \Delta\theta$ wavelengths in a direction parallel to that of the source. This results in a phase change of approximately $2\pi \Delta u \Delta\theta$ in the fringe component corresponding to the spacing $(u + \Delta u)$, of which points A_1 and B are an example. Since this phase change is linearly proportional to Δu , the variation of the visibility within the range $(u - d_\lambda)$ to $(u + d_\lambda)$ can be obtained by Fourier transformation of the correlator output with

respect to the pointing offset $\Delta\theta$. Thus, the changes in pointing induce variations in the fringe phase that are dependent on the spacing of the incoming rays within the antenna apertures, and this effect allows the information on the variation of the visibility to be retained.

The conclusion given above, that the scanning action of the antennas allows information on a range of visibility values to be retrieved, was first reached by Ekers and Rots (1979), using a mathematical analysis, as follows. Consider a pair of antennas with spacing (u_0, v_0) pointing in the direction (l_p, m_p) . As the pointing angle is varied, the effective intensity distribution over the field of interest is represented by $I(l, m)$ convolved with the *normalized* antenna beam $A_N(l, m)$. The observed fringe visibility is the Fourier transform with respect to u and v of $I(l, m)$ multiplied by the antenna response for the particular pointing:

$$\mathcal{V}(u_0, v_0, l_p, m_p) = \int \int A_N(l - l_p, m - m_p) I(l, m) e^{-j2\pi(u_0 l + v_0 m)} dl dm . \quad (11.12)$$

Assuming that the antenna beam is symmetrical, we can write Eq. (11.12) as

$$\mathcal{V}(u_0, v_0, l_p, m_p) = \int \int A_N(l_p - l, m_p - m) I(l, m) e^{-j2\pi(u_0 l + v_0 m)} dl dm , \quad (11.13)$$

which has the form of a two-dimensional convolution:

$$\mathcal{V}(u_0, v_0, l_p, m_p) = [I(l, m) e^{-j2\pi(u_0 l + v_0 m)}] * * A_N(l, m) . \quad (11.14)$$

Now we take the Fourier transform of \mathcal{V} with respect to u and v , which represents the full-field visibility data obtained by means of the ensemble of pointing angles used:

$$\begin{aligned} \mathcal{V}(u, v) &= \int \int [I(l, m) e^{-j2\pi(u_0 l + v_0 m)}] * * A_N(l, m) e^{j2\pi(ul + vm)} dl dm \\ &= [\mathcal{V}(u, v) * * {}^2\delta(u_0 - u, v_0 - v)] \bar{A}_N(u, v) . \end{aligned} \quad (11.15)$$

Here, $\bar{A}_N(u, v)$ is the Fourier transform of $A_N(l, m)$, that is, the autocorrelation of the field distribution over the aperture of a single antenna, referred to as the transfer function or spatial sensitivity function of the antenna. The two-dimensional delta function ${}^2\delta(u_0 - u, v_0 - v)$ is the Fourier transform of $e^{-j2\pi(u_0 l + v_0 m)}$. As the final step, Eq. (11.15) becomes

$$\mathcal{V}(u, v) = \mathcal{V}[(u_0 - u), (v_0 - v)] \bar{A}_N(u, v) . \quad (11.16)$$

The conclusion from Eq. (11.16) is that if one observes a field of dimensions equal to several beamwidths, obtains the visibility for a number of pointing directions, and then for each antenna pair takes the Fourier transform of the visibility with respect to the pointing direction, the result will be values of the visibility extended over an

area of the (u, v) plane as large as the support of the function $\overline{A}_N(u, v)$. For a circular reflector antenna of diameter d , $\overline{A}_N(u, v)$ is nonzero within a circle of diameter $2d$. Thus, if $\overline{A}_N(u, v)$ is known with sufficient accuracy, that is, the beam pattern is sufficiently well calibrated, the visibility can be obtained at the intermediate points required to provide the full-field image.

In the practical reduction of visibility data used in mosaicking, the Fourier transform with respect to pointing is usually not explicitly performed. The importance of the discussion above is that it shows that the information at the required spacing is present in the data if the antenna pointing is scanned with respect to the source, either as a continuous motion or as a series of discrete pointings. The reduction to obtain the intensity distribution is generally based on the use of nonlinear deconvolution algorithms.

Cornwell (1988) has pointed out that the angular spacing required between the pointing centers on the sky can be deduced from the sampling theorem of Fourier transforms (Sect. 5.2.1). A more general form of the theorem can be stated as follows: If a function $f(x)$ is nonzero only within an interval of width Δ in the x coordinate, then it is fully specified if its Fourier transform $F(s)$ is sampled at intervals no greater than Δ^{-1} in s . If the sampling is coarser than this, aliasing will occur, and the original function will not be reproducible from the samples. Here, we consider an antenna beam pointing toward a source that is wide enough to cover most of the reception pattern, that is, the main beam and major sidelobes. As we move the antenna beam to different pointing angles to cover the source, we are effectively sampling the convolution of the source and the antenna beam. The beam pattern is equal to the Fourier transform of the autocorrelation function of the field distribution over the antenna aperture. The field cuts off at the edges of the aperture, which is d_λ wavelengths wide. Thus, the autocorrelation function cuts off at a width $2d_\lambda$. The sampling theorem indicates that the interval between pointings Δl_p should not exceed $1/(2d_\lambda)$ in order to fully sample the source convolved with the beam. In practice, the antenna illumination function is likely to be tapered at the edge, so the autocorrelation function falls to low levels before it reaches the cutoff width $2d_\lambda$. Thus, if Δl_p slightly exceeds $1/(2d_\lambda)$, the error introduced may not be large.

11.5.1 Methods of Producing the Mosaic Image

The basic steps in the mosaicking method are:

1. Observe the visibility function for an appropriate series of pointing centers.
2. Reduce the data for each pointing center independently to produce a series of images, each covering approximately one antenna beam area.
3. Combine the beam-area images into the required full-field image.

In step 2, it is desirable also to deconvolve the synthesized beam response from each beam-area image to remove the effects of sidelobes in the response, and this can be

done using, for example, CLEAN or MEM. Use of these nonlinear algorithms can fill in some of the frequency components of the intensity that were omitted from the coverage of the antenna array. Cornwell (1988) and Cornwell et al. (1993) describe two procedures for mosaic imaging. The first of these, which they refer to as linear mosaicking, is essentially the three steps above with a least-mean-squares procedure for combination of the individual pointing images in step 3. Although a nonlinear deconvolution is used individually on each beam-area image, the combination of the images is a linear process. The second procedure, which differs in that the deconvolution is performed jointly, is referred to as nonlinear mosaicking and involves a nonlinear algorithm such as MEM. Unmeasured visibility data can best be estimated in the deconvolution process if the full field that is covered by the ensemble of pointing angles contributes simultaneously to the deconvolution, rather than by treating each primary beam area separately. The benefit of a joint deconvolution of the combined beam-area images is illustrated by consideration of an unresolved component of the intensity distribution located at the edge of a beam area, where it occurs in two or more individual beam images. Being at the beam edge where the response is changing rapidly, the amplitude of the component is more likely to be inaccurately determined, but such errors will tend to average out in the combined data. In the application to mosaicking, maximum entropy can be envisaged as the formation of an image that is consistent with all the visibility data for the various pointings, within the uncertainty resulting from the noise.

Cornwell (1988) discusses use of the MEM algorithm of Cornwell and Evans (1985) in mosaicking. This algorithm is briefly described in Sect. 11.2.1 [see Eq. (11.5)]. The procedure is essentially the same as in the application to a single-pointing image, except for a few more steps in determining χ^2 and its gradient. As in Eq. (11.4), χ^2 is the statistic that indicates the deviation of the model from the measured visibility values and is here expressed as

$$\chi^2 = \sum_p \sum_k \frac{|\mathcal{V}_{kp}^{\text{meas}} - \mathcal{V}_{kp}^{\text{model}}|^2}{\sigma_{kp}^2}, \quad (11.17)$$

where the subscripts k and p indicate the k th visibility value at the p th pointing position, and σ_{kp}^2 is the variance of the visibility. An initial model is required, and the procedure follows a series of steps described by Cornwell (1988):

1. For the first pointing center, multiply the current trial model with the antenna beam as pointed during the observation, and take the Fourier transform with respect to (l, m) to obtain the predicted visibility values.
2. Subtract the measured visibilities from the model visibilities to obtain a set of residual visibilities. Insert the residual visibilities into the accumulating χ^2 function of Eq. (11.17).
3. By Fourier transformation, convert the residual visibilities, weighted inversely as their variances, into an intensity distribution. Taper this distribution by multiplying it by the antenna beam pattern, and store in a data array of dimensions equal to the full MEM model.

4. Repeat steps 1–3 for each pointing. In step 2, add the value for χ^2 to those for the other pointings in this cycle. In step 3, add the residual intensity values into the data array. The accumulated values in this data array are used to obtain the gradient of χ^2 with respect to the MEM image.

The reason for the additional multiplication of the residual distribution by the beam function in step 3 is that it reduces unwanted responses from sidelobes of the primary beam that fall on adjacent pointing areas. It also weights the data with respect to the signal-to-noise ratio. Completion of the MEM procedure may require several tens of cycles through the steps given above to obtain convergence to the final image. To complete the process, smoothing with a two-dimensional Gaussian beam of width equal to the array resolution is recommended, to reduce the effects of variable resolution across the image.

A slightly different procedure for nonlinear mosaicking is described by Sault et al. (1996). In this case, the beam-area images are combined linearly without the individual deconvolution step, and then the final nonlinear deconvolution is applied to the combined image. In the linear combination, each pixel in the combined image is a weighted sum of the corresponding pixels in the individual beam-area images. As an example, Sault et al. also show results for a mosaic of the Small Magellanic Cloud made with the compact configuration of the Australia Telescope using 320 pointings. They demonstrate that the joint deconvolution used in nonlinear mosaicking is superior to the linear combination of the subfield images, even if these have been individually deconvolved. They also show the deconvolution using both their method and that described by Cornwell (1988) and conclude that the results are of comparable quality.

11.5.2 *Short-Baseline Measurements*

In imaging sources wider than the antenna beam, it is important to obtain visibility values at increments in u and v that are smaller than the diameter of an antenna. Data equivalent to an essentially continuous coverage in u and v can then be obtained by observing at various pointing positions as discussed above. The minimum spacing of two antennas is limited by mechanical considerations, and there is a gap or region of low sensitivity corresponding to a spacing of about half the minimum spacing between the centers of two antenna apertures. This is called the “short-spacing problem.”

This minimum spacing depends on the antenna design, but in general, unless the range of zenith angles is restricted, two antennas of diameter d cannot be spaced much closer than about $1.4d$, or perhaps $1.25d$ with special design. Otherwise, there is danger of mechanical collision, especially if there is a possibility that the antennas may not always be pointing in the same direction. Total-power observations with a single antenna will, in principle, provide spacings from zero to d/λ , but with some antennas, measurements at spatial frequencies greater than $\sim 0.5d/\lambda$ are unreliable

because the spatial sensitivity function of the antenna falls to low levels as a result of the tapered illumination of the reflector. Missing data at low (u, v) values result in broad negative sidelobes of the synthesized beam, such that the beam appears to be situated in a shallow bowl-shaped depression. This effect is most noticeable when the field to be imaged is wide enough that there are several empty (u, v) cells within the central area.

The transfer function $\bar{A}_N(u)$ is the autocorrelation function of the field distribution over the antenna aperture and depends on the particular design of the antenna, including the illumination pattern of the feed. The solid curve in Fig. 11.7 shows \bar{A}_N for a uniformly illuminated circular aperture, which can be regarded as an ideal case. Since there is usually some tapering in the illumination of a reflector antenna, in practice, \bar{A}_N will generally fall off somewhat more rapidly than the curves shown. The function \bar{A}_N in Fig. 11.7 is proportional to the common area of two overlapping circles of diameter d , and the abscissa is the distance between their centers. In three dimensions, this function is sometimes referred to as the Chinese hat function, and its properties are discussed by Bracewell (1995). The dashed curves in Fig. 11.7 show the relative spatial sensitivity for an interferometer using two uniformly illuminated circular apertures of diameter d . Curve 1 is for a spacing of $1.4d$ between the centers of the apertures; curve 2 is for a spacing of $1.25d$. If both total-power and interferometer data are obtained, it can be seen that the minimum sensitivity occurs for spacings of approximately half of the antenna spacing.

One solution to increasing the minimum sensitivity in the spatial frequency coverage is the addition of total-power measurements from a larger antenna [see, for example, Bajaja and van Albada (1979), Welch and Thornton (1985), Stanimirović (2002)]. Stanimirović considered the requirements for the use of single-antenna measurements of fringe visibility and concluded that the diameter of the antenna

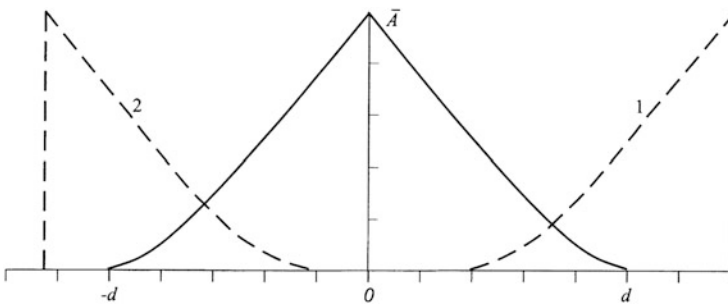


Fig. 11.7 The solid curve centered on the origin shows the spatial sensitivity function \bar{A}_N for a single antenna of diameter d . The curve corresponds to the case of uniform excitation over the aperture. This curve indicates the relative sensitivity to spatial frequencies for total-power observations with a single antenna. The dashed curves show the spatial sensitivity for two antennas of diameter d , with uniform aperture excitation, working as an interferometer. Curve 1 is for a spacing of $1.4d$ between the centers of the antennas, and curve 2 is for a spacing of $1.25d$. If the aperture illumination is tapered, the curves will fall off to low values more rapidly than is shown.

should be at least 1.5 times the spacing for which the visibility value is required. Note, however, that since the cost of an antenna scales approximately as $d^{2.7}$ (see Sect. 5.7.2.2), the expected cost of an antenna of diameter $1.5d$ is roughly 4.4 times that of an antenna of diameter d . The process of merging total power and interferometric data is sometimes called “feathering.”

Another possibility for covering the missing spatial frequencies is the use of one or more pairs of smaller antennas, say, $d/2$ in diameter, with spacing about $0.7d$. A pair of antennas of diameter $d/2$ have one-quarter of the area, and consequently one-quarter of the sensitivity to fine structure, of a pair of the standard antennas. Since the beam of the smaller antenna has four times the solid angle of a standard antenna, it will require one-quarter of the number of pointing directions, and the integration time for each one can be four times as long. Cornwell et al. (1993) present evidence that, for mosaicking, it is possible to obtain satisfactory performance with a homogeneous array, that is, one in which all antennas are the same size. This requires total-power observation as well as interferometry with some antennas spaced as closely as possible. The deconvolution steps in the data reduction help to fill in remaining (u, v) gaps.

At frequencies of several hundred gigahertz, where antenna beams are of minute-of-arc order, images of objects of order one degree in size require numbers of pointings in the range 10^2 – 10^4 . Any given pointing cannot be quickly repeated, so dependence on Earth rotation to fill in small gaps in the (u, v) coverage may not be practicable. Thus, arrays designed for mosaicking of large objects require good instantaneous (u, v) coverage. At such high frequencies, it is also desirable to avoid high zenith angles to minimize atmospheric effects.

An alternative to tracking discrete pointing centers is to sweep the beams over the area of sky under investigation in a raster scan motion. This technique has been referred to as “on-the-fly” mosaicking. It has several advantages:

- The uniformity of the (u, v) coverage for all points in the field is maximized, which results in uniformity of the synthesized beam across the resulting image and thereby simplifies the image processing.
- Each point in the field is observed many times in as rapid succession as possible, so some advantage can be taken of Earth rotation to fill in the (u, v) coverage.
- If total-power measurements are made, the scanning motion of the beam can be used to remove atmospheric effects in a similar way to the use of beam switching in large single-dish telescopes.
- Waste of observing time during moves of the antennas from one pointing center to another is eliminated.

With on-the-fly observing, the real-time integration at the correlator output must be somewhat less than the time taken for the beam to scan over any point in the field, and thus a large number of visibility data, each with a separate pointing position, are generated.

11.6 Multifrequency Synthesis

Making observations at several different radio frequencies is an effective way of improving the sampling of the visibility in the (u, v) plane. This technique is referred to as multifrequency synthesis, or bandwidth synthesis. Generally, the range of frequencies is about $\pm 15\%$ of the midrange value. Such a range can be very effective in filling in gaps in the coverage, and since it is not too large, major changes in the source structure with frequency are avoided [see, e.g., Conway et al. (1990)]. However, the variation of structure with frequency may be large enough to limit the dynamic range unless some steps are taken to mitigate it, as discussed here. The principal continuum radio emission mechanisms produce radio spectra that vary smoothly in frequency (see Fig. 1.1), and the intensity usually follows a power-law variation with frequency:

$$I(\nu) = I(\nu_0) \left(\frac{\nu}{\nu_0} \right)^\alpha, \quad (11.18)$$

where α is the spectral index, which varies with (l, m) . If the spectrum does not conform to a power law, then, in effect, we can write

$$\alpha = \frac{\nu}{I} \frac{\partial I}{\partial \nu}. \quad (11.19)$$

If the spectral index were a constant over the source, the spectral effects could be removed. Although this is generally not the case, the spectral effects are reduced by first correcting the data for a “mean” or “representative” spectral index for the overall structure to be imaged. Thus, from this point, α will represent the spectral index of the deviation of the intensity distribution from this first-order correction. Consider the case in which the intensity variation can be approximated by a linear term:

$$\begin{aligned} I(\nu) &= I(\nu_0) + \frac{\partial I}{\partial \nu} (\nu - \nu_0) \\ &= I(\nu_0) + \alpha I(\nu_0) \frac{(\nu - \nu_0)}{\nu} \\ &\simeq I(\nu_0) + \alpha I(\nu_0) \frac{(\nu - \nu_0)}{\nu_0}, \end{aligned} \quad (11.20)$$

where the reference frequency ν_0 is near the center of the range of frequencies used. Equation (11.20) is the sum of a single-frequency term and a spectral term. To determine the synthesized beam of an array working in the multifrequency mode, consider the response to a point source with a spectrum given by Eq. (11.20). The response to the single-frequency term can be obtained by taking the Fourier transform of the spatial transfer function. The transfer function has a delta function

of u and v for each visibility measurement. Each frequency used contributes a different set of delta functions. The response to the spectral term is obtained by multiplying the transfer function by $(\nu - \nu_0)/\nu_0$ and taking the Fourier transform. If we call the single-frequency and spectral responses b'_0 and b'_1 , respectively, the synthesized beam is equal to

$$b_0(l, m) = b'_0(l, m) + \alpha(l, m)b'_1(l, m) . \quad (11.21)$$

The first component is a conventional synthesized beam, and the second one is an unwanted artifact. The measured intensity distribution obtained as the Fourier transform of the measured visibilities is

$$I_0(l, m) = I(l, m) ** b'_0(l, m) + \alpha(l, m)I(l, m) ** b'_1(l, m) , \quad (11.22)$$

where $I(l, m)$ is the true intensity on the sky. Conway et al. (1990) and Sault and Wieringa (1994) have both developed deconvolution processes based on the CLEAN algorithm that deconvolve both b'_0 and b'_1 . In the method used by the first of these groups, components representing each one of the two beams were removed alternately. In the method used by the second group, each component removed represented both beams. These methods provide the distribution of both the source intensity and the spectral index as functions of frequency. Conway et al. also consider a logarithmic rather than a linear form of the frequency offsets from ν_0 . These analyses show that for a frequency spread of approximately $\pm 15\%$, the magnitude of the response resulting from the b'_1 component is typically 1% and can sometimes be ignored. Removing the b'_1 component reduces the spectral effects to $\sim 0.1\%$.

For other approaches and extensions to multifrequency synthesis, see Rau and Cornwell (2011) and Junklewitz et al. (2015).

11.7 Noncoplanar Baselines

In Sect. 3.1, it was shown that, except in the case of east–west linear arrays, the baselines of a synthesis array do not remain in a plane as the Earth rotates. It was also shown that for fields of view of small angular size [as given approximately by Eq. (3.12)], the Fourier transform relationship between visibility and intensity can be expressed satisfactorily in two dimensions. However, particularly for frequencies less than a few hundred megahertz, the small-field assumption does not always apply. At meter wavelengths, the primary beams of the antennas are wide, for example, $\sim 6^\circ$ for a 25-m-diameter antenna at a wavelength of 2 m. Also, the high density of strong sources on the sky at meter wavelengths requires that the full beam be imaged to avoid confusion. We now consider the case in which the condition in Eq. (3.12) ($\theta_f < \frac{1}{3}\sqrt{\theta_b}$) is not valid, so the two-dimensional solution should not be used. The following treatment follows those of Sramek and Schwab (1989) and

others as noted. We start with the exact result in Eq. (3.7), which is

$$\mathcal{V}(u, v, w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{A_N(l, m)I(l, m)}{\sqrt{1 - l^2 - m^2}} \times \exp \left\{ -j2\pi \left[ul + vm + w \left(\sqrt{1 - l^2 - m^2} - 1 \right) \right] \right\} dl dm . \quad (11.23)$$

Here, $\mathcal{V}(u, v, w)$ is the visibility as a function of spatial frequency in three dimensions, $A_N(l, m)$ is the normalized primary beam pattern of an antenna, and $I(l, m)$ is the two-dimensional intensity distribution to be imaged.

The next step is to rewrite Eq. (11.23) in the form of a three-dimensional Fourier transform, which involves the third direction cosine n defined with respect to the w axis. The phase of the visibility $\mathcal{V}(u, v, w)$ is measured relative to the visibility of a (hypothetical) source at the phase reference position for the observation. This introduces a factor $e^{j2\pi w}$ within the exponential term on the right side of Eq. (11.23), as noted in the text following Eq. (3.7). The corresponding phase shift is inserted by the fringe rotation discussed in Sect. 6.1.6. As a result of this factor, we use $n' = n - 1$ as the conjugate variable of w in order to obtain the three-dimensional Fourier transform. Functions of n' will be indicated by a prime. Thus, we rewrite Eq. (11.23):

$$\mathcal{V}(u, v, w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{A_N(l, m)I(l, m)}{\sqrt{1 - l^2 - m^2}} \delta \left(\sqrt{1 - l^2 - m^2} - n' - 1 \right) \times \exp \{ -j2\pi (ul + vm + wn') \} dl dm dn' . \quad (11.24)$$

The delta function $\delta(\sqrt{1 - l^2 - m^2} - n' - 1)$ is introduced to maintain the condition $n = \sqrt{1 - l^2 - m^2}$ and thereby to allow n' to be treated as an independent variable in the Fourier transformation. In a practical observation, \mathcal{V} is measured only at points at which the sampling function $W(u, v, w)$ is nonzero. The Fourier transform of the sampled visibility defines a three-dimensional intensity function I'_3 as follows:

$$I'_3(l, m, n') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(u, v, w) \mathcal{V}(u, v, w) e^{j2\pi(ul + vm + wn')} du dv dw . \quad (11.25)$$

This is the Fourier transform of the product of the two functions $W(u, v, w)$ and $\mathcal{V}(u, v, w)$, which by the convolution theorem is equal to the convolution of the Fourier transforms of the two functions. Thus,

$$I'_3(l, m, n') = \left\{ \frac{A_N(l, m)I(l, m) \delta \left(\sqrt{1 - l^2 - m^2} - n' - 1 \right)}{\sqrt{1 - l^2 - m^2}} \right\} \overline{W}(l, m, n') . \quad (11.26)$$

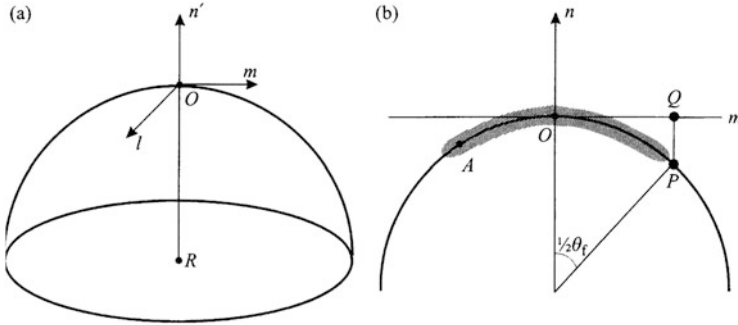


Fig. 11.8 (a) One hemisphere of the unit sphere in (l, m, n) coordinates. The point R is the origin of the (l, m, n) coordinates. O is the origin of the (l, m, n') coordinates, which is the phase reference point. (b) Section through the unit sphere in the (m, n) plane. The shaded area represents the extent of the function I_3 . A source at point A would not appear, or would be greatly attenuated, in a two-dimensional analysis in the (l, m) plane. The width of the three-dimensional “beam” in the n direction should be comparable to that in l and m , since the range of the sampling function in w is comparable to that in u and v if the observations cover a large range in hour angle. (In the superficially similar case in Fig. 3.5, the intensity function is not confined to the surface of the sphere because the measurements are all made in the $w' = 0$ plane).

Here, $\overline{W}'(l, m, n')$ is the Fourier transform of the three-dimensional sampling function $W(u, v, w)$, and the triple asterisk denotes three-dimensional convolution. Having determined the result of the Fourier transformation, we can now replace n' by $(n - 1)$, and Eq. (11.26) becomes

$$I_3(l, m, n) = \left\{ \frac{A_N(l, m) I(l, m) \delta(\sqrt{1 - l^2 - m^2} - n)}{\sqrt{1 - l^2 - m^2}} \right\} \overline{W}(l, m, n) \quad (11.27)$$

The expression in the braces on the right side of Eq. (11.27) is confined to the surface of the unit sphere $n = \sqrt{1 - l^2 - m^2}$, since the delta function is nonzero only on the sphere. The function \overline{W} with which it is convolved is the Fourier transform of the sampling function and is, in effect, a three-dimensional dirty beam. The convolution has the effect of spreading the expression so that I_3 has finite extent in the radial direction of the sphere. Figure 11.8a shows the unit sphere centered on the origin of (l, m, n) coordinates at R . The (l, m) plane in which the results of the conventional two-dimensional analysis lie is tangent to the unit sphere at O , at which point $n = 1$ and $n' = 0$. Note that since l , m , and n are direction cosines, the unit sphere in (l, m, n) is a mathematical concept, not a sphere in real space.

Several ways of obtaining an undistorted wide-field image are possible (Cornwell and Perley 1992).

1. *Three-Dimensional Transformation.* $I_3(l, m, n)$ can be deconvolved by means of a three-dimensional extension of the CLEAN algorithm. This is complicated

by the fact that the visibility is, in practice, not as well sampled in w as it is in u and v . From Fig. 3.4, the large values of w occur for large zenith angles of the target source. In Fig. 11.8b, the width of the angular field is θ_f . The transform must be computed over the range of (l, m) within this field, and over the range PQ in n . Cornwell and Perley (1992) suggest using a direct (rather than discrete) Fourier transform in the n to w transformation, since otherwise the poor sampling may result in serious sidelobes and aliasing. Thus, two-dimensional FFTs are performed in a series of planes normal to the n axis. The number of planes required is equal to PQ divided by the required sampling interval in the n direction. The range of measured visibility values has a width $2|w|_{\max}$ in the w direction, so, by the sampling theorem, the intensity function is fully specified in the n coordinate if it is sampled at intervals of $(2|w|_{\max})^{-1}$. The distance PQ is approximately equal to $\frac{1}{8}\theta_f^2 \approx \frac{1}{2}|l^2 + m^2|_{\max}$ [note that the angle $POQ = \theta_f/4$, and $(\theta_f/2)^2 = |l^2 + m^2|_{\max}$]. Thus, the number of planes in which the two-dimensional intensity must be calculated is $|l^2 + m^2|_{\max} |w|_{\max}$. [This result can also be obtained by taking the phase term in Eq. (3.8) that is omitted in going from three to two dimensions and sampling at the Nyquist interval of half a turn of phase.] The maximum possible value of w is D_{\max}/λ , where D_{\max} is the longest baseline in the array. If θ_f is limited by the beamwidth of antennas of diameter d , for which the angular distance from the beam center to the first null is $\sim \lambda/d$, the required number of planes is $\sim (\lambda/d)^2 \times D_{\max}/\lambda = \lambda D_{\max}/d^2$. Examples of images made using this method are given by Cornwell and Perley (1992).

2. *Polyhedron Imaging.* The area of the unit sphere for which the image is required can be divided into a number of subfields, which can be imaged individually using the small-field approximation. Each one is imaged in two dimensions onto a plane that is tangent to the unit sphere at a different point on the sphere. These tangent points are the phase centers for the individual subfields. For each subfield image, it is necessary to adjust both the visibility phases and the (u, v, w) coordinates of the whole database to the particular phase center. The subfields can be combined using methods similar to those used in mosaicking, including joint deconvolution. This approach has been referred to as *polyhedron imaging* because the various image planes form part of the surface of a polyhedron. Again, examples are given by Cornwell and Perley (1992).
3. *Combination of Snapshots.* In most synthesis arrays, the antennas are mounted on an area of approximately level ground and thus lie close to a plane at any given instant. In such cases, a long observation can be divided into a series of “snapshots,” for each of which the planar baseline condition applies individually. It should therefore be possible to make an image by combining a series of snapshot responses. Each snapshot represents the true intensity distribution convolved with a different dirty beam, since the (u, v) coverage changes progressively as the source moves across the sky. Ideally, deconvolution would thus require optimization of the intensity distribution using the snapshot responses in a combined manner rather than individually. It should be noted that the plane in which the baselines lie for any snapshot is, in general, not normal to

the direction of the target source. As a result, the angle at which points on the unit sphere in Fig. 11.8a are projected onto the (l, m) plane is not parallel to the n axis and varies with the position of the source on the sky. Positions of sources in the snapshot images suffer an offset in (l, m) that is zero at the phase center but that increases with distance from the phase center. Images should be corrected for this effect before being combined. Since the required correction varies with the hour angle of the source, in long observations, the effect can cause smearing of source details in the outer part of the field. Perley (1999b) discusses this effect and its correction. Bracewell (1984) has discussed a method similar to the combination of snapshots described above.

4. *Deconvolution with Variable Point-Source Response.* In some cases, the effect of two-dimensional Fourier transformation is principally distortion of the point-source response in the outer parts of the field, without serious attenuation of the response. Then a possible procedure is deconvolution using a point-source response (the dirty beam) that is varied over the field to match the calculated response (McClellan (1984)). This approach was used by Waldram and McGilchrist (1990) in analysis of a survey using the Cambridge Low-Frequency Synthesis Telescope, which operated at 151 MHz using Earth rotation and baselines that are offset from east–west by 3° . Point-source responses were computed for a grid of positions within the field, and the response for any particular position could then be obtained by interpolation. The principal requirement was to obtain accurate positions and flux densities for sources identified in images obtained by two-dimensional transformation. Fitting the appropriate theoretical beam response for each source position allowed distortion of the beam, including any position offset, to be accounted for. The procedure was relatively inexpensive in computer time.
5. *W-Projection.* *W*-projection (Cornwell et al. 2008) is a more efficient method of handling the problem of noncoplanar baselines. This problem occurs when the width of the synthesized field is sufficiently large that the w term in the exact visibility equation [(3.7) and (11.23)] cannot be neglected. In *w*-projection, we start by rewriting the visibility equation, Eq. (11.23), as

$$\mathcal{V}(u, v, w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{A_N(l, m) I(l, m)}{\sqrt{1 - l^2 - m^2}} G(l, m, w) e^{[-j2\pi(u l + v m)]} dl dm, \quad (11.28)$$

where

$$G(l, m, w) = e^{-j2\pi w(\sqrt{1 - l^2 - m^2} - 1)}, \quad (11.29)$$

so that the w dependence is contained within $G(l, m, w)$, and the other parts of Eq. (11.28) represent $\mathcal{V}(u, v, w = 0)$. If $\mathcal{G}(u, v, w)$ is the Fourier transform of $G(l, m, w)$ with respect to (u, v) and (l, m) , Eq. (11.28) can be written as a two-dimensional convolution in (u, v) ,

$$\mathcal{V}(u, v, w) = \mathcal{V}(u, v, w = 0) ** \mathcal{G}(u, v, w). \quad (11.30)$$

Again, we can visualize (u, v, w) space with the u and v axes in a horizontal plane with w increasing vertically upward. The measured visibility values are located within a block of (u, v, w) space of dimensions limited by the longest antenna spacings and the geometry of the observations. Generally, the observations are designed to optimize the uniformity of sampling of the visibility in the u and v dimensions, but the sampling in w is usually relatively sparse. The procedure in w -projection is to project the three-dimensional visibility data onto the $(u, v, w = 0)$ plane, from which a two-dimensional Fourier transform provides an image in l and m . The $(u, v, w = 0)$ plane is parallel to the tangent plane on the celestial sphere at the field center and thus represents data for which the ray paths from a source at the field center to the corresponding pair of antennas are of equal length. Data for which w is nonzero are those for which the ray paths differ in length by w wavelengths. To use such data to obtain visibility in the $(u, v, w = 0)$ plane, it is necessary to account for the additional path length to one antenna of each pair. In propagating the extra distance in space, the radiation from a point is spread by diffraction, so a single (u, v, w) point is spread into a diffraction pattern at $w = 0$. This spread of the pattern results from the width of the convolution function $\mathcal{G}(u, v, w)$ in Eq. (11.30) and is approximately proportional to $|w|$.

If we use the approximation $\sqrt{1 - l^2 - m^2} \approx 1 - (l^2 + m^2)/2$, Eq. (11.29) becomes

$$G(l, m, w) \approx e^{-j\pi w(l^2 + m^2)}. \quad (11.31)$$

Fourier transformation then gives

$$\mathcal{G}(u, v, w) \approx \frac{j}{w} e^{-j\pi(u^2 + v^2)/w}. \quad (11.32)$$

The visibility $\mathcal{V}(u, v, w)$ is entirely determined by $\mathcal{V}(u, v, w = 0)$ through convolution with \mathcal{G} . Thus, $\mathcal{V}(u, v, w = 0)$ contains all the data that are required to provide an accurate image, limited only by the synthesized (dirty) beam. Nothing essential to the image is lost in the transition from three dimensions to two. The same convolution function \mathcal{G} applies to projection in both directions, i.e., from $(u, v, w = 0)$ to (u, v, w) and vice versa. Note that the convolving function is different for each (u, v, w) data point. Cornwell et al. (2008) point out that this convolutional relationship between a two-dimensional and a three-dimensional one is due to the fact that the original brightness is confined to the two-dimensional surface of the celestial sphere. They also discuss the result in terms of the diffraction of the electric field over the w -coordinate space.

The w -projection imaging procedure is as follows. First, the visibility data are gridded in (u, v, w) and then projected onto the $(u, v, w = 0)$ plane. In

the projection, the data are spread in (u, v) space by the convolution,¹ and thus regridding in the $(u, v, w = 0)$ plane is required. A two-dimensional Fourier transform then provides the dirty image, from which the dirty beam must then be deconvolved by the CLEAN algorithm or some alternate procedure (see Sect. 11.1). CLEAN requires numerous transpositions of data between the visibility and image domains. In going from the model image to visibility, a two-dimensional transform provides $\mathcal{V}(u, v, w = 0)$, from which projection gives values at (u, v, w) points required for comparison with the observations. For the regridding steps, convolution with a spheroidal or other gridding function is required. Since convolution is commutative and associative, it can be computationally efficient to convolve the spheroidal function with the projection function \mathcal{G} and thus store the combined convolution functions for use with each (u, v, w) grid point. Convolution of \mathcal{G} with the spheroidal function has the additional benefit of damping the behavior of \mathcal{G} as $w \rightarrow 0$.

Cornwell et al. (2008) also provide details of a simulated example of wide-field imaging using w -projection. They compare the results with the method of image-plane facets (similar to polyhedron imaging), and also uvw -space facets (similar to mosaicking), which projects the (u, v) space, rather than image space, onto tangent plane facets. Hitherto, the facets methods has been perhaps the most widely used procedure for wide-field imaging. Cornwell et al. conclude that, with regard to computing load, the facets method is roughly competitive with w -projection for images of low dynamic range but that w -projection is superior when high sensitivity and dynamic range are required.

A variation of w -projection imaging, which is computationally less expensive, is called w -stacking (Offringa et al. 2014).

11.8 Some Special Techniques of Image Analysis

11.8.1 *Use of CLEAN and Self-Calibration with Spectral Line Data*

A procedure that has been found to provide accurate separation of the continuum from the line features involves use of the deconvolving algorithm CLEAN (van Gorkom and Ekers 1989). However, if CLEAN is applied individually to the images for the different channels, errors in the CLEAN process appear as differences from channel to channel and may be confused with true spectral features. Such errors can be avoided by subtracting the continuum before applying CLEAN to the line data. First, CLEAN is applied to an average of the continuum-only channels,

¹Cornwell et al. (2008) point out that, interestingly, this spreading in (u, v) space shows that, in general, a single antenna pair responds to a range of spatial frequencies, except when $w = 0$ (i.e., when the baseline is normal to the direction of incidence of the wavefront.)

and the visibility components removed from these channels are also removed from the visibility data for the channels containing line features. When the CLEAN process is terminated, the residuals are also removed from the line data. The resulting line-channel images, which should then contain only line data, can be deconvolved individually. Note that since absorption of the continuum may occur in the line frequency channels, images of line-minus-continuum may contain negative as well as positive intensity features. Thus, algorithms such as MEM that depend on positivity of the intensity may not be easily applicable in such cases.

In applying self-calibration to eliminate phase errors in spectral line data, it can generally be assumed that phase and amplitude differences between channels vary only very little with time and are removed by the bandpass calibration. This is true for both atmospheric and instrumental effects. Thus, the strongest spectral feature in the field under investigation can be used to determine the phase-calibration solution, which is then applied to all channels.

11.8.2 *A-Projection*

Observations of the most distant Universe, which require removal of the emissions from the foreground, require observations at the highest precision and correspondingly accurate calibration of instrumental effects. Calibration of the responses of the individual antennas includes correcting for DD gains in the deconvolution of images, as discussed by Bhatnagar et al. (2008), Smirnov (2011a,b), and others. DD gains² include instrumental and atmospheric effects that affect the pointing and polarization of the antenna responses. Correction for DD effects includes taking account of the rotation of the antennas relative to the sky that results from altazimuth tracking. The DD effects for each antenna can be represented in a 2×2 Jones matrix, a separate one for each pixel of the image. For each cross-correlated antenna pair, the signal product is represented by the outer product of the two Jones matrices, which provides a 4×4 Mueller matrix for each pixel. The diagonal elements of the Mueller matrix represent the four principal products of the two cross-polarization terms of each antenna, for either linear or circular polarization. The off-diagonal terms are small and result from errors in the cross-polarization adjustment and from leakage. These terms must be included if accuracy better than $\sim 1\%$ is required in the image. This procedure has been referred to as the narrowband *A*-projection algorithm, in which *A* refers to the elements of $A_{i,j}$ that are the complex convolution of the aperture illumination patterns of antennas *i* and *j*. The details of the cross products depend upon the details of the particular array, and Bhatnagar et al. (2008) consider the case for the VLA, in which shading by the feed legs is one of the factors represented by the off-axis terms. The derivation of the image from observations involves an

²Direction-independent gains include, for example, the gain of the receiver system and are generally much simpler to correct for.

iterative χ^2 minimalization in which χ represents the difference between observed visibility and the visibility of a model that is developed. Calculation of gradients of χ^2 can be used as an aid in the minimalization.

Bhatnagar et al. (2013) expand these concepts to cover a wider frequency bandwidth and develop an *A*-projection algorithm that includes variation of the model parameters as a function of frequency. A wide bandwidth ratio (e.g., 2 : 1) improves the sensitivity of the observations but requires careful consideration since in the outer parts of the beam, the response varies rapidly with frequency. On wideband, wide-field imaging, see also Rau and Cornwell (2011). An extension of the *A*-projection technique called fast holographic deconvolution, which is particularly useful for very wide field-of-view observations, has been developed by Sullivan et al. (2012).

11.8.3 *Peeling*

Synchrotron emission from radio sources usually becomes stronger as the frequency is reduced, and hence, the density of strong sources on the sky generally increases with decreasing frequency. At low frequencies, it is therefore often important to image the whole antenna beam to avoid source confusion resulting from aliasing. Also, the gain of the main beam of a reflector antenna decreases with decreasing frequency, and if phased arrays of dipoles are used, they have to be very large to maintain high gain. As a result, sources in the sidelobes may not be as effectively suppressed relative to a source in the main beam, as is possible at higher frequencies. In the data analysis, unwanted responses from strong sources with known positions can be subtracted. In this process, known as peeling (Noordam 2004; van der Tol et al. 2007), the response to such sources, down to the lowest calibrated level of the sidelobes, is removed. This usually starts with the strongest source in the field and then the second strongest, and so on. The removal can be done in the visibility domain. A procedure of this type is essential in the measurement of the weakest sources and the Epoch of Reionization signatures (see Sect. 10.7.2). Some further discussion of peeling can be found in Bhatnagar et al. (2008), Mitchell et al. (2008), and Bernardi et al. (2011).

11.8.4 *Low-Frequency Imaging*

In addition to source confusion, a complication of the wide-field imaging is the variation of ionospheric effects over the field of view (see Sect. 14.1). The excess path length in the ionosphere is proportional to ν^{-2} , so the resulting phase change is proportional to ν^{-1} . The term *isoplanatic patch* is used to denote an area of the sky over which the variation in the path length for an incoming wave is small

compared with the observing wavelength. At centimeter and shorter wavelengths, the beams of reflector antennas used in synthesis arrays are generally smaller than the isoplanatic patch (see Table 14.1). Thus, the effect of an irregularity in the ionosphere (or troposphere) is constant over the beam and can be corrected by a single phase adjustment for each antenna, for example, by self-calibration. However, at meter wavelengths, the size of the antenna beam may be several times that of the ionospheric isoplanatic patch. In observations with the VLA in New Mexico, Erickson (1999) estimated that the size of the isoplanatic patch at 74 MHz is $\sim 3^\circ$ – 4° , whereas the beamwidth of a 25-m-diameter antenna at the same frequency is $\sim 13^\circ$. Later, low-frequency instruments have used phased arrays, which enable much smaller beams to be formed. These include LOFAR, covering 15–80 MHz and 110–240 MHz (de Vos et al. 2009); the Murchison Widefield Array, covering 80–300 MHz (Lonsdale et al. 2009); and the Long Wavelength Array, covering 10–88 MHz (Ellingson et al. 2009).

Although, at meter wavelengths, arrays of dipoles or similar elements are more generally used than parabolic reflectors, some early measurements using the 25-m-diameter antennas of the VLA by Kassim et al. (1993) are of interest. These include simultaneous measurements of a number of strong sources at 74 and 330 MHz, using a phase reference procedure to calibrate the phases at the lower frequency. At 74 MHz, the phase fluctuations are dominated by the ionosphere, and rates of phase change were found to be as high as one degree per second. These precluded calibration by the usual methods. However, at 330 MHz, the rates of phase change were slow enough to allow imaging of strong sources. The resulting 330-MHz phases were scaled to 74 MHz and used to remove the ionospheric component from the 74-MHz data that were recorded simultaneously. The procedure used for obtaining images at 74 MHz was essentially as follows:

1. Simultaneous observations of a strong source were made at 74 and 330 MHz, with periodic observations of a calibrator at 330 MHz.
2. An image of the target source was made at 330 MHz using the standard techniques (i.e., use of a calibrator as at centimeter wavelengths). This was used as a starting model for self-calibration of the 330-MHz data.
3. The self-calibration provided phase calibration for each antenna at 330 MHz. These values were then scaled to 74 MHz and used to remove the ionospheric variations from the 74 MHz data, the ionospheric phase changes being inversely proportional to frequency.
4. The instrumental phases at 330 and 74 MHz were different at each antenna as a result of dissimilar cable lengths, etc. To calibrate these differences, an unresolved calibrator was observed at both 330 and 74 MHz. The ionospheric variations could be removed from the 74-MHz calibrator phases using the scheme in step 3.
5. The 74-MHz image of the target source was made from the calibrated phase data. Self-calibration of the 74-MHz data was used to remove residual phase drifts, and for this, the 330-MHz image provided a suitable starting model.

For the strongest sources, for which it was possible to obtain a good signal-to-noise ratio in an averaging time of no more than 10 s, self-calibration at 74 MHz was sufficient in most cases. Although only eight VLA antennas were equipped for operation at 74 MHz, images with dynamic range of better than 20 dB were obtained for several sources. The problem of noncoplanar baselines did not arise in these measurements because the sources were compact enough for satisfactory two-dimensional imaging.

11.8.5 *Lensclean*

There are many cases in which the image of a quasar or radio galaxy is distorted by the gravitational field of a galaxy, following the discovery of this phenomenon by Walsh et al. (1979). The line of sight from the lens source intersects, or passes very close to, the galaxy. In some cases, the gravitational lensing results in multiple images of a single point-source quasar, and in other cases, extended structure is involved: see, for example, Narayan and Wallington (1992). In studies of gravitational lensing, the structure of the gravitational field is of major astrophysical importance. The term *lensclean* has been used to denote a method of analysis, including several variations of the original algorithm, that allows the lensing field to be determined by synthesis imaging. The basis of these methods is analogous to self-calibration, in which the image is sufficiently overdetermined by the visibility measurements that it is possible to determine also the complex gains of the antennas. In *lensclean*, it is the pattern of the gravitational field that is determined. An additional complication is that points in the source of the radiation can each contribute to more than one point in the synthesized image.

The original *lensclean* procedure (Kochanek and Narayan 1992) is based on an adaptation of the CLEAN algorithm. The basic principle can be described as follows. Consider the case in which the source that is imaged by the lens contains extended structure. An initial model for the lens is chosen. Each point in the source contributes to multiple points in the image, and this procedure from the source to the image is defined by the lens model. For any point in the source, the intensity in the image should ideally be the same at each point at which it appears, since the imaging involves only geometric bending of the radiation from the source, as in an optical system. Suppose that the j th source pixel contributes to n_j image pixels. In practice, the intensity of these pixels in the image is not equal because of defects in the lens model and noise in the image. The best estimate of the intensity of the pixel in the source is the mean intensity of the corresponding pixels in the image. Thus, one can subtract components from the image in the manner of CLEAN and build up an image of the source. For each source pixel for which $n_j > 1$, the mean-squared deviation of the intensity of the corresponding image pixels from the mean intensity of the $n_j > 1$ image pixels, σ_j^2 , is calculated. For a good lens model, the mean value of σ_j^2 over the pixels in the source should be no greater than the variance of the noise

in the image, σ_{noise}^2 . If the number of degrees of freedom in the source image is taken to be equal to the number of pixels, then the statistical measure of the quality of the lens model is $\chi^2 = \sum(\sigma_j^2/\sigma_{\text{noise}}^2)$, where the sum is taken over the j source pixels. The lens parameters can thus be varied to minimize χ^2 . In practice, the procedure is more complicated than indicated by the description above. Modifications are included to take account of the finite resolution of the image, which has the effect of spreading the contribution of each source pixel over a number of image pixels. Also, for any unresolved structure in the source, the intensity of the corresponding structure in the image depends on the magnification of the lens.

Ellithorpe et al. (1996) introduced a *visibility lensclean* procedure in which the CLEAN components are removed from the ungridded visibility values under the constraints of a lens model. The squared deviations of the measured visibility from a model are used to determine a χ^2 statistic. The quality of the fit is judged from the variance of the measured visibility, and the number of degrees of freedom is $2N_{\text{vis}} - 3N_{\text{src}} - N_{\text{lens}}$. Here, N_{vis} is the number of visibility measurements (which each have two degrees of freedom), N_{src} is the number of independent CLEAN components in the source model (three degrees of freedom, from position and amplitude), and N_{lens} is the number of parameters in the lens model. Ellithorpe et al. compared results of the original lensclean with visibility lensclean and found the best results from the latter, with a further improvement if a self-calibration step is added. The use of the MEM algorithm as an alternative to CLEAN has also been investigated (Wallington et al. 1994).

11.8.6 Compressed Sensing

Compressed sensing, also known as compressive sensing, compressive sampling, or sparse sampling, is a widely used signal processing technique generally employed to reduce the size of data sets, e.g., images, without loss of information. Sampling at the Nyquist interval provides the most general and complete representation of an image. However, if an image is sparse, i.e., it is mostly blank with isolated components or can be represented by a small number of basis functions such as wavelets, then it is possible to compress or reduce the image size far below that required for Nyquist sampling. The theory of compressed sensing has formal requirements such as sparsity and incoherent sampling. The latter requirement in interferometric imaging corresponds to random sampling in the (u, v) plane. Under such conditions, an image can be derived exactly with very high probability from a sparse set of visibility measurements. These conditions are not perfectly met in radio interferometry. Nonetheless, much can be learned from compressed sensing techniques [see Li et al. (2011a,b)].

In the application to interferometric imaging, the method is formulated in a way to obtain an accurate image from an incompletely sampled (u, v) plane data set. The degree of success of the method depends on the signal-to-noise ratio in the

(u, v) plane and the amount of information that can be supplied to constrain the image solution while being consistent with the (u, v) plane measurements. The simplest of these constraints are nonnegativity, compactness, and smoothness in the image plane. In other words, the method improves as the amount of a priori information available increases. For applications to radio interferometric data and specific algorithms, see, for example, Wiaux et al. (2009), Wenger et al. (2010), Li et al. (2011), Hardy (2013), Carrillo et al. (2012), Garsden et al. (2015), and Dabbech et al. (2015). For a general introduction to the field of compressed sensing as a signal-processing tool, see Candès and Wakin (2008), and for its application to image construction, see Candès et al. (2006a,b). Compressed sensing is widely used in medical imaging [e.g., Lustig et al. (2008)].

For a simplified overview of the method and some of its key concepts, consider a linear vector equation that can be expressed as

$$\mathbf{V} = \mathbf{A}\mathbf{X}, \quad (11.33)$$

where \mathbf{V} represents visibility, \mathbf{X} represents brightness in the image, and \mathbf{A} is the operator that derives the visibility from the parameters of the image, i.e., the Fourier transform kernels.

Important quantities in the image-restoration process are the L_p norms defined as

$$L_p \equiv \|\mathbf{X}\|_p = \left[\sum_{n=1}^N |X_n|^p \right]^{1/p}, \quad p > 0, \quad (11.34)$$

where X_n are the elements of \mathbf{X} . For $p = 0$, a pseudo norm L_0 can be written as

$$L_0 = \|\mathbf{X}\|_0 = \sum_{n=1}^N |X_n|^0, \quad (11.35)$$

with the understanding that $0^0 \equiv 1$. In our context, L_0 is the number of cells in the image with nonzero amplitude. Suppose that a point source is observed. In the absence of measurement noise, the normalized moduli of the visibilities will all be unity. Minimizing L_0 with the image constraint of Eq. (11.33) will lead to the recovery of a source distribution of a delta function. Note that the principal image solution is proportional to the dirty image. L_0 minimization tends to remove its high sidelobe response.

Unfortunately, exploration of the L_0 norm is computationally very intensive. In two of the foundational papers in compressed sensing, Candès and Tao (2006) and Donoho (2006) showed that under fairly general conditions, the L_1 norm, defined as

$$L_1 = \|\mathbf{X}\|_1 = \sum_{n=1}^N |X_n|, \quad (11.36)$$

is a suitable proxy for the L_0 norm, and it is more easily computed. The L_1 norm is the total flux density for sources of positive brightness. Virtually all work in compressed sensing is based on L_1 minimization. In the interferometry case, the solution process can be described as

$$\text{minimize } \|\mathbf{X}\|_1, \quad \text{subject to } \|\mathbf{V} - \mathbf{A}\mathbf{X}\|_2^2 < \epsilon, \quad (11.37)$$

where ϵ is the noise threshold based on the measurements, and $\|\mathbf{V} - \mathbf{A}\mathbf{X}\|_2^2$ is the squared sum of the visibility residuals or the goodness of fit. An equivalent optimization can be written as

$$\text{minimize } \left\{ \|\mathbf{V} - \mathbf{A}\mathbf{X}\|_2^2 + \Lambda \|\mathbf{X}\|_1 \right\}, \quad (11.38)$$

where Λ is a regularization parameter that determines the relative importance of minimizing L_1 and the measurement residuals. In statistics, this approach is called the *least absolute shrinkage selection operator*, or LASSO, developed by Tibshirani (1996).

Another commonly used constraint is based on total variation (TV), often computed for two-dimensional images as

$$TV = \sum_{i,j} \left[(X_{i-1,j} - X_{i,j})^2 + (X_{i,j+1} - X_{i,j})^2 \right]^{1/2}. \quad (11.39)$$

TV is also known as the L_1 norm for adjacent pixel differences. Minimizing TV minimizes the gradients and favors smoother images. TV minimization can be added to Eq. (11.38) with another Λ term. Note that MEM imposes a similar smoothness constraint (see Sect. 11.2.1). A nonnegative constraint can also be added. Collectively, the application of these constraints is known as regularization.

The potential for recovering source structure finer than the diffraction limit has been investigated by Honma et al. (2014). An example of this application is shown in Fig. 11.9. The possibilities of successful superresolution improve as the image plane becomes more sparse, i.e., a “near black” image [see Starck et al. (2002)].

Another variation on the above approach, more useful for extended source distributions, is to represent the image by a set of basis functions such as wavelets [see Starck and Murtagh (1994) and Starck et al. (1994)]. If the representation is sparse in such a basis space, L_1 minimization gives good results [e.g., Li et al. (2011a,b) and Garsden et al. (2015)].

The most efficient and reliable methods of producing images from visibility data are the subject of continuing development. As with the advent of MEM methods, it is often desirable for researchers to present multiple reconstructions of images that are consistent with their (u, v) plane measurements.

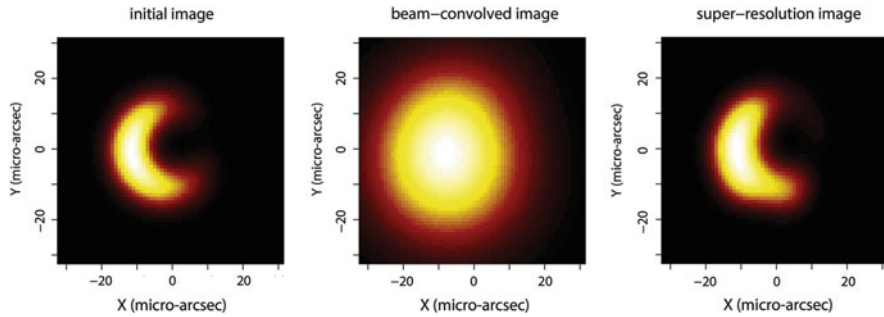


Fig. 11.9 Reconstructed images of a simulated black hole shadow source observed with a six-element Event Horizon Telescope Array at the position of M87. **(left)** Simulated image, **(middle)** CLEANed image with resolution equal to that of the dirty beam, and **(right)** image reconstructed with compressed sensing regularization methods. It may be difficult to apply techniques tested on simulated data. From M. Honma et al. (2014), by permission of and © Oxford University Press.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Further Reading

- Roberts, J.A., Ed., *Indirect Imaging*, Cambridge Univ. Press, Cambridge, UK (1984)
 Taylor, G.B., Carilli, C.L., and Perley, R.A., Eds., *Synthesis Imaging in Radio Astronomy II*, Astron. Soc. Pacific Conf. Ser., **180** (1999)
 van Schooneveld, C., Ed., *Image Formation from Coherence Functions in Astronomy*, Reidel, Dordrecht (1979)

References

- Ables, J.G., Maximum Entropy Spectral Analysis, *Astron. Astrophys. Suppl.*, **15**, 383–393 (1974)
 Bajaja, E., and van Albada, G.D., Complementing Aperture Synthesis Radio Data by Short Spacing Components from Single Dish Observation, *Astron. Astrophys.*, **75**, 251–254 (1979)
 Baldwin, J.E., and Warner, P.J., Phaseless Aperture Synthesis, *Mon. Not. R. Astron. Soc.*, **182**, 411–422 (1978)

- Baldwin, J.E., and Warner, P.J., Fundamental Aspects of Aperture Synthesis with Limited or No Phase Information, in *Image Formation from Coherence Functions in Astronomy*, van Schooneveld, C., Ed., Reidel, Dordrecht (1979), pp. 67–82
- Bates, R.H.T., Contributions to the Theory of Intensity Interferometry, *Mon. Not. R. Astron. Soc.*, **142**, 413–428 (1969)
- Bates, R.H.T., Uniqueness of Solutions to Two-Dimensional Fourier Phase Problems for Localized and Positive Images, *Comp. Vision, Graphics, Image Process.*, **25**, 205–217 (1984)
- Bernardi, G., Mitchell, D.A., Ord, S.M., Greenhill, L.J., Pindor, B., Wayth, R.B., and Wyithe, J.S.B., Subtraction of Point Sources from Interferometric Radio Images Through an Algebraic Modelling Scheme, *Mon. Not. R. Astron. Soc.*, **413**, 411–422 (2011)
- Bhatnagar, S., Cornwell, T.J., Golap, K., and Uson, J.M., Correcting Direction-Dependent Gains in the Deconvolution of Radio Images, *Astron. Astrophys.*, **487**, 419–429 (2008)
- Bhatnagar, S., Rau, U., and Golap, K., Wide-Field Wide-Band Interferometric Imaging: A WB A-Projection and Hybrid Algorithms, *Astrophys. J.*, **770**:91 (9pp) (2013)
- Bracewell, R.N., Inversion of Nonplanar Visibilities, in *Indirect Imaging*, Roberts, J.A., Ed., Cambridge Univ. Press, Cambridge, UK (1984), pp. 177–183
- Bracewell, R.N., *Two-Dimensional Imaging*, Prentice-Hall, Englewood Cliffs, NJ (1995)
- Braun, R., Understanding Synthesis Imaging Dynamic Range, *Astron. Astrophys.*, **551**, A91 (26 pp.) (2013)
- Briggs, D.S., “High Fidelity Deconvolution of Moderately Resolved Sources,” Ph.D. thesis, New Mexico Institute of Mining and Technology (1995). <http://www.aoc.nrao.edu/dissertations/dbriggs>
- Bruck, Y.M., and Sodin, L.G., On the Ambiguity of the Image Reconstruction Problem, *Opt. Commun.*, **30**, 304–308 (1979)
- Bryan, R.K., and Skilling, J., Deconvolution by Maximum Entropy, as Illustrated by Application to the Jet of M87, *Mon. Not. R. Astron. Soc.*, **191**, 69–79 (1980)
- Candès, E.M., Romberg, J., and Tao, T., Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information, *IEEE Trans. Inf. Theory*, **52**, 489–509 (2006a)
- Candès, E.M., Romberg, J., and Tao, T., Stable Signal Recovery from Incomplete and Inaccurate Measurements, *Comm. Pure Appl. Math.*, **59**, 1207–1223 (2006b)
- Candès, E.M., and Tao, T., Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies?, *IEEE Trans. Inf. Theory*, **52**, 5406–5425 (2006)
- Candès, E.J., and Wakin, M.B., An Introduction to Compressive Sampling, *IEEE Signal Proc. Mag.*, **25**, 21–30 (2008)
- Carrillo, R.E., McEwen, J.D., and Wiaux, Y., Sparsity Averaging Reweighted Analysis (SARA): A Novel Algorithm for Radio-Interferometric Imaging, *Mon. Not. R. Astron. Soc.*, **426**, 1223–1234 (2012)
- Clark, B.G., An Efficient Implementation of the Algorithm “CLEAN,” *Astron. Astrophys.*, **89**, 377–378 (1980)
- Clark, B.G., Large Field Mapping, in *Synthesis Mapping*, Proc. NRAO Workshop No. 5, Thompson, A.R., and D’Addario, L.R., Eds., National Radio Astronomy Observatory, Green Bank, WV (1982)
- Conway, J.E., Cornwell, T.J., and Wilkinson, P.N., Multi-Frequency Synthesis: A New Technique in Radio Interferometric Imaging, *Mon. Not. R. Astr. Soc.*, **246**, 490–509 (1990)
- Cornwell, T.J., Self Calibration, in *Synthesis Mapping*, Proc. of NRAO Workshop No. 5, Thompson, A.R., and D’Addario, L.R., Eds., National Radio Astronomy Observatory, Green Bank, WV (1982)
- Cornwell, T.J., A Method of Stabilizing the Clean Algorithm, *Astron. Astrophys.*, **121**, 281–285 (1983)
- Cornwell, T.J., Radio-Interferometric Imaging of Very Large Objects, *Astron. Astrophys.*, **202**, 316–321 (1988)
- Cornwell, T.J., The Applications of Closure Phase to Astronomical Imaging, *Science*, **245**, 263–269 (1989)

- Cornwell, T.J., Imaging Concepts, in *Very Long Baseline Interferometry and the VLBA*, Zensus, J.A., Diamond, P.J., and Napier, P.J., Eds., Astron. Soc. Pacific Conf. Ser., **82**, 39–56 (1995)
- Cornwell, T.J., Multiscale CLEAN Deconvolution of Radio Synthesis Images, *IEEE J. Selected Topics in Signal Proc.*, **2**, 793–801 (2008)
- Cornwell, T.J., Braun, R., and Briggs, D.S., Deconvolution, in *Synthesis Imaging in Radio Astronomy II*, Taylor, G.B., Carilli, C.L., and Perley, R.A., Eds., Astron. Soc. Pacific Conf. Ser., **180**, 151–170 (1999)
- Cornwell, T.J., and Evans, K.F., A Simple Maximum Entropy Deconvolution Algorithm, *Astron. Astrophys.*, **143**, 77–83 (1985)
- Cornwell, T.J., Golap, K., and Bhatnagar, S., The Non-Coplanar Baselines Effect in Radio Astronomy. The W Projection Algorithm, *IEEE J. Selected Topics Signal Proc.*, **2**, 647–657 (2008)
- Cornwell, T.J., Holdaway, M.A., and Uson, J.M., Radio-Interferometric Imaging of Very Large Objects: Implications for Array Design, *Astron. Astrophys.*, **271**, 697–713 (1993)
- Cornwell, T.J., and Perley, R.A., Radio-Interferometric Imaging of Very Large Fields, *Astron. Astrophys.*, **261**, 353–364 (1992)
- Cornwell, T.J., and Wilkinson, P.N., A New Method for Making Maps with Unstable Radio Interferometers, *Mon. Not. R. Astron. Soc.*, **196**, 1067–1086 (1981)
- Cotton, W.D., A Method of Mapping Compact Structure in Radio Sources Using VLBI Observations, *Astron. J.*, **84**, 1122–1128 (1979)
- Dabbech, A., Ferrari, C., Mary, D., Slezak, E., Smirnov, O., and Kenyon, J.S., MORESANE: MODEL REconstruction by Synthesis-ANalysis Estimators: A Sparse Deconvolution Algorithm for Radio Interferometric Imaging, *Astron. Astrophys.*, **576**, A7 (16pp) (2015)
- de Vos, M., Gunst, A.W., and Nijboer, R., The LOFAR Telescope: System Architecture and Signal Processing, *Proc. IEEE*, **97**, 1431–1437 (2009)
- Donoho, D.L., Compressed Sensing, *IEEE Trans. Inf. Theory*, **52**, 1289–1306 (2006)
- Ekers, R.D., and Rots, A.H., Short Spacing Synthesis from a Primary Beam Scanned Interferometer, in *Image Formation from Coherence Functions in Astronomy*, van Schoonveld, C., Ed., Reidel, Dordrecht (1979), pp. 61–63
- Ellingson, S.W., Clarke, T.E., Cohen, A., Craig, J., Kassim, N.E., Pihlström, Y., Rickard, L.J., and Taylor, G.B., The Long Wavelength Array, *Proc. IEEE*, **97**, 1421–1430 (2009)
- Ellithorpe, J.D., Kochanek, C.S., and Hewitt, J.N., Visibility Lensclean and the Reliability of Deconvolved Radio Images, *Astrophys. J.*, **464**, 556–567 (1996)
- Erickson, W.C., Long Wavelength Interferometry, in *Synthesis Imaging in Radio Astronomy II*, Taylor, G.B., Carilli, C.L., and Perley, R.A., Eds., Astron. Soc. Pacific Conf. Ser., **180**, 601–612 (1999)
- Fienu, J.R., Reconstruction of an Object from the Modulus of Its Fourier Transform, *Opt. Lett.*, **3**, 27–29 (1978)
- Fort, D.N., and Yee, H.K.C., A Method of Obtaining Brightness Distributions from Long Baseline Interferometry, *Astron. Astrophys.*, **50**, 19–22 (1976)
- Frieden, B.R., Restoring with Maximum Likelihood and Maximum Entropy, *J. Opt. Soc. Am.*, **62**, 511–518 (1972)
- Garsden, H., Girard, J.N., Starck, J.L., Corbel, S., Tasse, C., Woiselle, A., McKean, J.P., van Amelsfoort, A.S., Anderson, J., Avruch, I.M., and 71 coauthors, LOFAR Sparse Image Reconstruction, *Astron. Astrophys.*, **575**, A90 (18 pp) (2015)
- Gull, S.F., and Daniell, G.J., Image Reconstruction from Incomplete and Noisy Data, *Nature*, **272**, 686–690 (1978)
- Gull, S.F., and Daniell, G.J., The Maximum Entropy Method, in *Image Formation from Coherence Functions in Astronomy*, van Schoonveld, C., Ed., Reidel, Dordrecht (1979), pp. 219–225
- Hardy, S.J., Direct Deconvolution of Radio Synthesis Images Using L_1 Minimization, *Astron. Astrophys.*, **557**, A134 (10pp) (2013)
- Högbom, J.A., Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines, *Astron. Astrophys. Suppl.*, **15**, 417–426 (1974)

- Högbom, J.A., The Introduction of A Priori Knowledge in Certain Processing Algorithms, in *Image Formation from Coherence Functions in Astronomy*, van Schooneveld, C., Ed., Reidel, Dordrecht (1979), pp. 237–239
- Honma, M., Akiyama, K., Uemura, M., and Ikeda, S., Super-Resolution Imaging with Radio Interferometry Using Sparse Modeling, *Publ. Astron. Soc. Japan*, **66**, 95 (14pp) (2014)
- Jaynes, E.T., Prior Probabilities, *IEEE Trans. Syst. Sci. Cyb.*, **SSC-4**, 227–241 (1968)
- Jaynes, E.T., On the Rationale of Maximum-Entropy Methods, *Proc. IEEE*, **70**, 939–952 (1982)
- Junklewitz, H., Bell, M.R., and Enßelin, T., A New Approach to Multifrequency Synthesis in Radio Interferometry, *Astron. Astrophys.*, **581**, A59 (11pp) (2015)
- Junklewitz, H., Bell, M.R., Selig, M., and Enßelin, T., RESOLVE: A New Algorithm for Aperture Synthesis Imaging of Extended Emission in Radio Astronomy, *Astron. Astrophys.*, **586**, A76 (21pp) (2016)
- Kassim, N.E., Perley, R.A., Erickson, W.C., and Dwarakanath, K.S., Subarcminute Resolution Imaging of Radio Sources at 74 MHz with the Very Large Array, *Astron. J.*, **106**, 2218–2228 (1993)
- Kochanek, C.S., and Narayan, R., Lensclean: An Algorithm for Inverting Extended, Gravitationally Lensed Images with Application to the Radio Ring Lens PKS 1830-211, *Astrophys. J.*, **401**, 461–473 (1992)
- Lawson, C.L., and Hanson, R.J., *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ (1974)
- Li, F., Brown, S., Cornwell, T.J., and de Hoog, F., The Application of Compressive Sampling to Radio Astronomy. 2. Faraday Rotation Measure Synthesis, *Astron. Astrophys.*, **531**, A126 (8pp) (2011a)
- Li, F., Cornwell, T.J., and de Hoog, F., The Application of Compressive Sampling to Radio Astronomy. 1. Deconvolution, *Astron. Astrophys.*, **528**, A31 (10pp) (2011b)
- Lonsdale, C.J., Cappallo, R.J., Morales, M.F., Briggs, F.H., Benkevitch, L., Bowman, J.D., Bunton, J.D., Burns, S., Corey, B.E., deSouza, L., and 38 coauthors, The Murchison Widefield Array: Design Overview, *Proc. IEEE*, **97**, 1497–1506 (2009)
- Lonsdale, C.J., Doeleman, S.S., Capallo, R.J., Hewitt, J.N., and Whitney, A.R., Exploring the Performance of Large- N Radio Astronomical Arrays, in *Radio Telescopes*, Butcher, H.R., Ed., Proc. SPIE, **4015**, 126–134 (2000)
- Lustig, M., Donoho, D.L., Santos, J.M., and Pauly, J.M., Compressed Sensing MRI, *IEEE Signal Proc. Mag.*, **25**, 72–82 (2008)
- McClellan, D.J., A Simple Expansion Method for Wide-Field Mapping, in *Indirect Imaging*, Roberts, J.A., Ed., Cambridge Univ. Press, Cambridge, UK (1984), pp. 185–191
- Mitchell, D.A., Greenhill, L.J., Wayth, R.B., Sault, R.J., Lonsdale, C.J., Capallo, R.J., Morales, M.F., and Ord, S.M., Real-Time Calibration of the Murchison Widefield Array, *IEEE J. Selected Topics Signal Proc.*, **2**, 707–717 (2008)
- Napier, P.J., The Brightness Temperature Distributions Defined by a Measured Intensity Interferogram, *NZ J. Sci.*, **15**, 342–355 (1972)
- Napier, P.J., Bagri, D.S., Clark, B.G., Rogers, A.E.E., Romney, J.D., Thompson, A.R., and Walker, R.C., The Very Long Baseline Array, *Proc. IEEE*, **82**, 658–672 (1994)
- Napier, P.J., and Bates, R.H.T., Inferring Phase Information from Modulus Information in Two-Dimensional Aperture Synthesis, *Astron. Astrophys. Suppl.*, **15**, 427–430 (1974)
- Napier, P.J., Thompson, A.R., and Ekers, R.D., The Very Large Array: Design and Performance of a Modern Synthesis Radio Telescope, *Proc. IEEE*, **71**, 1295–1320 (1983)
- Narayan, R., and Nityananda, R., Maximum Entropy—Flexibility vs. Fundamentalism, in *Indirect Imaging*, Roberts, J.A., Ed., Cambridge Univ. Press, Cambridge, UK (1984), pp. 281–290
- Narayan, R., and Wallington, S., Introduction to Basic Concepts of Gravitational Lensing, in *Gravitational Lenses*, Kayser, R., Schramm, T., and Nieser, L., Eds., Springer-Verlag, Berlin (1992), pp. 12–26
- Nityananda, R., and Narayan, R., Maximum Entropy Image Reconstruction—A Practical Non-Information-Theoretic Approach, *J. Astrophys. Astron.*, **3**, 419–450 (1982)

- Noordam, J.E., LOFAR Calibration Challenges, in *Ground-Based Telescopes*, Oschmann, J.M., Jr., Ed., Proc. SPIE, **5489**, 817–825 (2004)
- Offringa, A.R., McKinley, B., Hurley-Walker, N., Briggs, F.H., Wayth, R.B., Kaplan, D.L., Bell, M.E., Feng, L., Neben, A.R., Hughes, J.D., and 43 coauthors, WSCLEAN: An Implementation of a Fast, Generic Wide-Field Imager for Radio Astronomy, *Mon. Not. R. Astron. Soc.*, **444**, 606–619 (2014)
- Pearson, T.J., and Readhead, A.C.S., Image Formation by Self-Calibration in Radio Astronomy, *Ann. Rev. Astron. Astrophys.*, **22**, 97–130 (1984)
- Perley, R.A., High Dynamic Range Imaging, in *Synthesis Imaging in Radio Astronomy*, Perley, R.A., Schwab, F.R., and Bridle, A.H., Eds., Astron. Soc. Pacific Conf. Ser., **6**, 287–313 (1989)
- Perley, R.A., High Dynamic Range Imaging, in *Synthesis Imaging in Radio Astronomy II*, Taylor, G.B., Carilli, C.L., and Perley, R.A., Eds., Astron. Soc. Pacific Conf. Ser., **180**, 275–299 (1999a)
- Perley, R.A., Imaging with Non-Coplanar Arrays, in *Synthesis Imaging in Radio Astronomy II*, Taylor, G.B., Carilli, C.L., and Perley, R.A., Eds., Astron. Soc. Pacific Conf. Ser., **180**, 383–400 (1999b)
- Ponsonby, J.E.B., An Entropy Measure for Partially Polarized Radiation and Its Application to Estimating Radio Sky Polarization Distributions from Incomplete “Aperture Synthesis” Data by the Maximum Entropy Method, *Mon. Not. R. Astron. Soc.*, **163**, 369–380 (1973)
- Rau, U., and Cornwell, T.J., A Multi-Scale Multi-Frequency Deconvolution Algorithm for Synthesis Imaging in Radio Astronomy, *Astron. Astrophys.*, **532**, A71 (17pp) (2011)
- Readhead, A.C.S., Walker, R.C., Pearson, T.J., and Cohen, M.H., Mapping Radio Sources with Uncalibrated Visibility Data, *Nature*, **285**, 137–140 (1980)
- Readhead, A.C.S., and Wilkinson, P.N., The Mapping of Compact Radio Sources from VLBI Data, *Astrophys. J.*, **223**, 25–36 (1978)
- Rogers, A.E.E., Methods of Using Closure Phases in Radio Aperture Synthesis, *Soc. Photo-Opt. Inst. Eng.*, **231**, 10–17 (1980)
- Rogers, A.E.E., Hinteregger, H.F., Whitney, A.R., Counselman, C.C., Shapiro, I.I., Wittels, J.J., Klemperer, W.K., Warnock, W.W., Clark, T.A., Hutton, L.K., and four coauthors, The Structure of Radio Sources 3C273B and 3C84 Deduced from the “Closure” Phases and Visibility Amplitudes Observed with Three-Element Interferometers, *Astrophys. J.*, **193**, 293–301 (1974)
- Sault, R.J., Staveland-Smith, L., and Brouw, W.N., An Approach to Interferometric Mosaicing, *Astron. Astrophys. Suppl.*, **120**, 375–384 (1996)
- Sault, R.J., and Wieringa, M.H., Multi-Frequency Synthesis Techniques in Radio Interferometric Imaging, *Astron. Astrophys. Suppl.*, **108**, 585–594 (1994)
- Schwab, F.R., Adaptive Calibration of Radio Interferometer Data, *Soc. Photo-Opt. Inst. Eng.*, **231**, 18–24 (1980)
- Schwab, F.R., Relaxing the Isoplanatism Assumption in Self Calibration: Applications to Low-Frequency Radio Astronomy, *Astron. J.*, **89**, 1076–1081 (1984)
- Schwarz, U.J., Mathematical-Statistical Description of the Iterative Beam Removing Technique (Method CLEAN), *Astron. Astrophys.*, **65**, 345–356 (1978)
- Schwarz, U.J., The Method “CLEAN”—Use, Misuse, and Variations, in *Image Formation from Coherence Functions in Astronomy*, van Schooneveld, C., Ed., Reidel, Dordrecht (1979), pp. 261–275
- Skilling, J., and Bryan, R.K., Maximum Entropy Image Reconstruction: General Algorithm, *Mon. Not. R. Astron. Soc.*, **211**, 111–124 (1984)
- Smirnov, O.M., Revisiting the Radio Interferometer Measurement Equation. I. A Full-Sky Jones Formalism, *Astron. Astrophys.*, **527**, A106 (2011a)
- Smirnov, O.M., Revisiting the Radio Interferometer Measurement Equation. II. Calibration and Direction-Dependent Effects, *Astron. Astrophys.*, **527**, A107 (2011b)
- Sramek, R.A., and Schwab, F.R., Imaging, in *Synthesis Imaging in Radio Astronomy*, Perley, R.A., Schwab, F.R., and Bridle, A.H., Eds., Astron. Soc. Pacific Conf. Ser., **6**, 117–138 (1989)

- Stanimirović, S., Short Spacings Correction from the Single-Dish Perspective, in *Single-Dish Radio Astronomy: Techniques and Applications*, Stanimirović, S., et al., Eds., Astron. Soc. Pacific Conf. Ser., **278**, 375–396 (2002)
- Starck, J.-L., Bijaoui, A., Lopez, B., and Perrier, C., Image Reconstruction by the Wavelet Transform Applied to Aperture Synthesis, *Astron. Astrophys.*, **283**, 349–360 (1994)
- Starck, J.-L., and Murtagh, F., Image Restoration with Noise Suppression Using the Wavelet Transform, *Astron. Astrophys.*, **228**, 342–348 (1994)
- Starck, J.-L., Pantin, E., and Murtagh, F., Deconvolution in Astronomy: A Review, *Publ. Astron. Soc. Pacific*, **114**, 1051–1069 (2002)
- Subrahmanya, C.R., An Optimum Deconvolution Method, in *Image Formation from Coherence Functions in Astronomy*, van Schooneveld, C., Ed., Reidel, Dordrecht (1979), pp. 287–290
- Sullivan, I.S., Morales, M.F., Hazelton, B.J., Arcus, W., Barnes, D., Bernardi, G., Briggs, F.H., Bowman, J.D., Bunton, J.D., Cappallo, R.J., and 41 coauthors, Fast Holographic Deconvolution: A New Technique for Precision Radio Interferometry, *Astrophys. J.*, **759**:17 (6pp) (2012)
- Thomasson, P., MERLIN, *Quart. J. R. Astron. Soc.*, **27**, 413–431 (1986)
- Tibshirani, R., Regression Shrinkage and Selection via the Lasso, *J. R. Statist. Soc. B*, **58**, 267–288 (1996)
- van der Tol, S., Jeffs, B.D., and van der Veen, A.-J., Self-Calibration for the LOFAR Radio Astronomical Array, *IEEE Trans. Signal Proc.*, **55**, 4497–4510 (2007)
- van Gorkom, J.H., and Ekers, R.D., Spectral Line Imaging II: Calibration and Analysis, in *Synthesis Imaging in Radio Astronomy*, Perley, R.A., Schwab, F.R., and Bridle, A.H., Eds., Astron. Soc. Pacific Conf. Ser., **6**, 341–353 (1989)
- Wakker, B.P., and Schwarz, J.J., The Multi-Resolution Clean and Its Application to the Short-Spacing Problem in Interferometry, *Astron. Astrophys.*, **200**, 312–322 (1988)
- Waldram, E.M., and McGilchrist, M.M., Beam-Sets: A New Approach to the Problem of Wide-Field Mapping with Non-Coplanar Baselines, *Mon. Not. R. Astron. Soc.*, **245**, 532–541 (1990)
- Wallington, S., Narayan, R., and Kochanek, C.S., Gravitational Lens Inversion Using the Maximum Entropy Method, *Astrophys. J.*, **426**, 60–73 (1994)
- Walsh, D., Carswell, R.F., and Weymann, R.J., 0957+561A,B: Twin Quasistellar Objects or Gravitational Lens?, *Nature*, **279**, 381–384 (1979)
- Welch, W.J., and Thornton, D.D., An Introduction to Millimeter and Submillimeter Interferometry and a Summary of the Hat Creek System, in *Int. Symp. Millimeter and Submillimeter Wave Radio Astronomy*, International Scientific Radio Union, Institut de Radio Astronomie Millimétrique, Granada, Spain (1985), pp. 53–64
- Wenger, S., Magnor, M., Pihlström, Y., Bhatnagar, S., and Rau, U., SparseRI: A Compressed Sensing Framework for Aperture Synthesis Imaging in Radio Astronomy, *Publ. Astron. Soc. Pacific*, **122**, 1367–1374 (2010)
- Wernecke, S.J., Two-Dimensional Maximum Entropy Reconstruction of Radio Brightness, *Radio Sci.*, **12**, 831–844 (1977)
- Wernecke, S.J., and D’Addario, L.R., Maximum Entropy Image Reconstruction, *IEEE Trans. Comput.*, **C-26**, 351–364 (1977)
- Wiaux, Y., Jacques, L., Puy, G., Scaife, A.M.M., and Vanderghynst, P., Compressed Sensing Imaging Techniques for Radio Interferometry, *Mon. Not. R. Astron. Soc.*, **395**, 1733–1742 (2009)